

STIC-Biotech/ChemLib

86128

Fr m: Bunner, Bridget  
Sent: Thursday, February 06, 2003 11:54 AM  
To: STIC-Biotech/ChemLib  
Subject: sequence search

Hi! I'd like a sequence search performed for case 09/762,594:

1. the nucleic acid sequence of SEQ ID NO: 2

Thanks!

Bridget Bunner

Art Unit 1647  
CM1-10D12  
(703) 305-7148  
mailbox 10B19

Point of Contact:  
Susan Hanley  
Technical Info. Specialist  
CM1 6B05 Tel: 305-4053

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 21:30:21 ; Search time 330 Seconds  
(without alignments)  
9956.562 Million cell updates/sec

Title: US-09-762-594-2  
Perfect score: 1459  
Sequence: 1 gaattcgccgcgtgcac.....catgaagttaattctctt 1459

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description
Result No.	Score	Query Match	Length	ID	
1	1459	100.0	1459	21	AAZ57038
2	1079.6	74.0	3399	22	AAK52051
3	1073.2	73.6	1481	21	AAZ87207
4	1045.4	71.7	2350	22	AAK53035
5	634.8	43.5	2714	21	AAA93116
6	293.2	20.1	420	21	AAZ7363
7	288.2	18.4	278	22	AAK53762
8	222.4	15.2	291	21	AAZ77466
9	204.4	14.0	934	24	ABK24346
					PBR-associated pro
					Human polynucleoti
					Human NTAP CDNA cl
					Human polynucleoti
					Human secreted pro
					Human ORFX ORF2918
					Murine transport a
					Human ORFX ORF3021
					DNA encoding human

10	142.6	9.8	873	22	AAH72906	Human cervical can
11	129	8.8	4384	22	AAK77371	Human immune/haema
12	113.4	7.8	527	24	ABK24345	DNA encoding human
13	110.6	7.6	554	22	ABA62955	Human foetal liver
14	110.6	7.6	554	22	AAK11364	Human brain expres
15	110.6	7.6	554	22	AAK37154	Human bone marrow
16	110.6	7.6	554	22	AAI17988	Probe #11664 used
17	110.6	7.6	554	22	AAI42978	Probe #11664 used
18	110.6	7.6	554	24	ABK11145	Human genome-deriv
19	109.2	7.5	306	22	ABA76496	Human foetal liver
20	109.2	7.5	306	22	AAK25143	Human brain expres
21	109.2	7.5	306	22	AAK51149	Human bone marrow
22	109.2	7.5	306	22	AAI28147	Probe #18080 for g
23	109.2	7.5	306	22	AAI57184	Probe #25870 used
24	109.2	7.5	306	24	ABK24667	Human genome-deriv
25	100.8	6.9	1495	23	ABL24635	Drosophila melanog
26	91.8	6.3	535	23	ABV55196	Human prostate exp
27	85.6	5.9	573	22	AAH71226	Human cervical can
28	78.4	5.4	486	22	AAK75507	Human cervical can
29	77.2	5.3	684	24	ABQ89793	Polyglutamine trac
30	75.2	5.2	4000	18	AAK91902	Human prostate exp
31	74.2	5.1	3489	21	AAA30290	Mannose-1-phosphat
32	74.2	5.1	3489	22	AAK82901	Kaposi's sarcoma-a
33	74.2	5.1	3489	24	ABA93487	Nucleotide sequenc
34	74.2	5.1	32207	20	AAV73805	Kaposi's sarcoma-a
35	74.2	5.1	137507	19	AAV19941	KSHV LUR DNA (nucl
36	74	5.1	1527	23	AAK68546	DNA encoding novel
37	74	5.1	1620	23	AAK68547	DNA encoding novel
38	74	5.1	2178	23	AAK86106	DNA encoding novel
39	73.8	5.1	49999	20	AAZ23895	Murine LOBO homolo
40	72.6	5.0	795	19	AAV55830	FLGA insert stabl
41	72.6	5.0	799	19	AAV55831	Nucleotide sequenc
42	72.6	5.0	1926	21	AAK50254	Epstein Barr virus
43	72.6	5.0	1926	22	AAK82902	EBV tethering prot
44	72.6	5.0	2580	21	AAV75454	Nucleotide sequenc
45	72.6	5.0	2580	24	AAI64275	Epstein-Barr virus

ALIGNMENTS

RESULT 1  
AAZ57038  
ID AAZ57038 standard; cDNA; 1459 BP.  
XX AC  
AC AAZ57038;  
XX DT 19-MAY-2000 (first entry)  
XX DE PBR-associated protein (PAP)7 encoding cDNA.  
XX DE PBR-associated protein (PAP)7 encoding cDNA.  
XX KW Peripheral-type benzodiazepine receptor; PBR; PBR-associated protein;  
XX KW PAP; cell proliferation; cancer; cell death; cytostatic; neuroprotective;  
XX KW immunomodulator; antiinfectility; cerebroprotective; atherosclerosis;  
XX KW Niemann-Pick C; tumour; Alzheimer's disease; developmental disorder;  
XX KW cholesterol; multiple sclerosis; stress; neurodegenerative disorder;  
XX KW immune disorder; stroke; PAP7; ss.  
XX OS Mus sp.  
XX FH Key  
XX FT CDS  
XX FT Location/Qualifiers  
XX FT 8..1399  
XX FT /\*tag= a  
XX FT /product= "PAP7"

WO200009549-A2.

24-FEB-2000.

11-AUG-1999; 99WO-US18507.

11-AUG-1998; 98US-0096048.

(GEO ) UNIV GEORGETOWN MEDICAL CENT.

Papadopoulos V, Li H;

WPI: 2000-224278/19.

DR P-PSDB: AAY67500.

Novel peripheral-type benzodiazepine receptor associated proteins used for the regulation of the peripheral-type benzodiazepine receptor

Claim 3; Page 68; 71pp; English.

The invention provides isolated peripheral-type benzodiazepine receptor (PBR) associated proteins (PAPs) and nucleic acids encoding the PAP proteins. The PAP polynucleotides are a source of primers and probes for detection, isolation and amplification. PAP ligands or substrates or antibodies can be labeled and used to detect PAPs, in the diagnosis or prognosis of disease associated with increased cell proliferation, such as cancer, or reduced cell death. The diagnostic methods of the invention can be predictive of diseases involving PBR including gallstones, atherosclerosis, Niemann-Pick C, Sirtosterolemia, Dystrophy, tumor proliferation, Schnyder's corneal crystalline dystrophy, brain disorders including Alzheimer's disease, cholesterol metabolism, Tellurium toxicity, Smith-Lemli-Opitz syndrome, myelinization, developmental abnormalities, demyelination, Charcot-Marie tooth disease, Pelizaeus-Werzbacher disease, Multiple sclerosis, and SLA. The methods may also be useful in prophylactic treatments, or in screening for compounds effective in prophylactic treatment. The PAPs may be used to identify inhibitors or activators which allows the identification of drugs or agents which modulate PBR activity. Inhibitors of PAP may be used in the treatment or amelioration of conditions such as stress and stroke, cancer, neurodegenerative disorders, developmental disorders, infertility and immune disorders. The present sequence represents a cDNA encoding a PAP7 polypeptide.

SQ Sequence 1459 BP; 455 A; 298 C; 403 G; 303 T; 0 other;

Query Match 100.0%; Score 1459; DB 21; Length 1459;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCGCGCTGCACCTAAAGTTGAGTTGCTACTGTAGTGACCCGTGTGAAGGT 60

DB 1 GAATTCGGCGCGCGCTGCACCTAAAGTTGAGTTGCTACTGTAGTGACCCGTGTGAAGGT 60

QY 61 AGTTTATTTTAAATCAACTTTCATTTGCAAACTAGTAAAGATGCCAAGCCCTTCA 120

DB 61 AGTTTATTTTAAATCAACTTTCATTTGCAAACTAGTAAAGATGCCAAGCCCTTCA 120

QY 121 TCCAACTTATGAAGAAAACCTGAAGTTGCTGGCACTGCATAGCAAGTTCTTTTGGGCC 180

DB 121 TCCAACTTATGAAGAAAACCTGAAGTTGCTGGCACTGCATAGCAAGTTCTTTTGGGCC 180

QY 181 ATATACCCAGACAGCTCCCTCCTGAGTTGATTTCTTGTGTTGGGGAATGATAGAG 240

DB 181 ATATACCCAGACAGCTCCCTCCTGAGTTGATTTCTTGTGTTGGGGAATGATAGAG 240

QY 241 GAGAGAAATGGGAGCTTGGGAAACATGTCGAAGGAGATGCCATGGTAGACTTTGTGAA 300

DB 241 GAGAGAAATGGGAGCTTGGGAAACATGTCGAAGGAGATGCCATGGTAGACTTTGTGAA 300

QY 301 GCTTCTTAATAAGTGTGTCTCTCTCTCGGCATATGTTGCTCCACACAGAAATAGAAA 360

DB 301 GCTTCTTAATAAGTGTGTCTCTCTCTCGGCATATGTTGCTCCACACAGAAATAGAAA 360

QY 361 GGAAGAAGAGAGAAAGAAAGCGGAGGAGCGAAGCGATGAGAGAGAGA 420

DB 361 GGAAGAAGAGAGAAAGAAAGCGGAGGAGCGAAGCGATGAGAGAGAGA 420

QY 421 ACGAGAGCGGCTGCAAAAGAGAGAGAAAGCGGAGGAGCGAAGCGATGAGAGAGAGA 480

DB 421 ACGAGAGCGGCTGCAAAAGAGAGAGAAAGCGGAGGAGCGAAGCGATGAGAGAGAGA 480

QY 481 ACGGAGAGGAGAGAGAGCGCGGATAGAGAAAGAGAGGCTTCGGCTGGAACACAAA 540

DB 481 ACGGAGAGGAGAGAGAGCGCGGATAGAGAAAGAGAGGCTTCGGCTGGAACACAAA 540

QY 541 GCAGCAGATATGCGAGCTTTAAACTCCAGACTGCCGTGCAATTCCAGCAGTATGCAGC 600

DB 541 GCAGCAGATATGCGAGCTTTAAACTCCAGACTGCCGTGCAATTCCAGCAGTATGCAGC 600

QY 601 CCAGCAGTATCCAGGGAACACGAAACACGAGATTCCTATCCGCCAGCTGCAGGAGCA 660

DB 601 CCAGCAGTATCCAGGGAACACGAAACACGAGATTCCTATCCGCCAGCTGCAGGAGCA 660

QY 661 GCACCTATCAGCAGTATAAACACCAGGAGCAACACCCTGCACAAACACGAGCAGC 720

DB 661 GCACCTATCAGCAGTATAAACACCAGGAGCAACACCCTGCACAAACACGAGCAGC 720

QY 721 ATTACAGAAACAGCAAGTAGTAGTGGCTGGGCATCATTTGCTGCATCATCAAGGT 780

DB 721 ATTACAGAAACAGCAAGTAGTAGTGGCTGGGCATCATTTGCTGCATCATCAAGGT 780

QY 781 GAACACAGCTGGAGCAAGTATACACTGCTCACTTAATGACAGGCCAAACCCACACTGA 840

DB 781 GAACACAGCTGGAGCAAGTATACACTGCTCACTTAATGACAGGCCAAACCCACACTGA 840

QY 841 AAATTCGAAAAGTCCCTTGAGCCAGAAAGCTGCAGAAAGCCCTTGGAAAATGGACCAA 900

DB 841 AAATTCGAAAAGTCCCTTGAGCCAGAAAGCTGCAGAAAGCCCTTGGAAAATGGACCAA 900

QY 901 AGACTCTCTTCAGTGTGAGTTCAGCTCCATCATGTTGGACAGACCAACAAAGACTT 960

DB 901 AGACTCTCTTCAGTGTGAGTTCAGCTCCATCATGTTGGACAGACCAACAAAGACTT 960

QY 961 TAAAGAGAAGATTGGCAGGATGCAGATTCTGTGATTACAGTACGTCGAGGAGAGTGGT 1020

DB 961 TAAAGAGAAGATTGGCAGGATGCAGATTCTGTGATTACAGTACGTCGAGGAGAGTGGT 1020

QY 1021 CACCGTCGAGTCCCGACTCATGAGAAAGATCATACCTATTTTGGGAATTTGCCACAGA 1080

DB 1021 CACCGTCGAGTCCCGACTCATGAGAAAGATCATACCTATTTTGGGAATTTGCCACAGA 1080

QY 1081 CAGTTATCACATTGGGTTTGGGTTTATTTGAATGGACAGACTCTCCAAATGCTGCTGT 1140

DB 1081 CAGTTATCACATTGGGTTTGGGTTTATTTGAATGGACAGACTCTCCAAATGCTGCTGT 1140

QY 1141 CAGTGTGATGTCAGTGTGAGTCCAGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200

DB 1141 CAGTGTGATGTCAGTGTGAGTCCAGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200

QY 1201 TGAAGAAAAGCAAAAAGAACGCCAACAGCCCTCTGCTGGATGAGATTGTACCTGTGTA 1260

DB 1201 TGAAGAAAAGCAAAAAGAACGCCAACAGCCCTCTGCTGGATGAGATTGTACCTGTGTA 1260

QY 1261 CCGCGGGAGCTGTACGAGGAGGAGTATATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320

DB 1261 CCGCGGGAGCTGTACGAGGAGGAGTATATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320

QY 1321 CTATCTCTCAAGTTTGAATTAATCTCTCTGTGTGAGGTCCTCAAGTCCGCTACTACAG 1380

DB 1321 CTATCTCTCAAGTTTGAATTAATCTCTCTGTGTGAGGTCCTCAAGTCCGCTACTACAG 1380

QY 1381 AGTCTATTATAGTAGAGTGTCTGTTCAGGTCGCGGAGTCCAGGTTTGTAGCACAAC 1440

DB 1381 AGTCTATTATAGTAGAGTGTCTGTTCAGGTCGCGGAGTCCAGGTTTGTAGCACAAC 1440

QY 1441 ATGAGCTTTAATTTCCCTTT 1459

DB 1441 ATGAGCTTTAATTTCCCTTT 1459

RESULT 2

AAK52051

ID AAK52051 standard: cDNA; 3399 BP.

XX





Db 970 TACAGTGGCCGAGGAGAAAGTGGTCACCTGTTCCGAGTACCCACCACCATGATGAAGAAGGATCATATA 1029  
Qy 1057 CCTATTTTGGGAATTTGCCACAGACAGTATGACATTTGGGTTTGGGTTTATTTTGAATG 1116  
Db 1030 TCTCTTTTGGGAATTTGCCACAGACAAATATGACATTTGGGTTTGGGTTTATTTTGAATG 1089  
Qy 1117 GACAGACTCTCAATGCTGTGTCAGTGTGCATGTCAAGTCCAGTCCAGTCCAGGAGGGA 1176  
Db 1090 GACAGACTCTCAACACTGCTGTGTCAGGTCATGTCTAGTCCAGTCCAGGATGACAGCA 1149  
Qy 1177 GGAGGAGGAAGAAATGTCACCTGTGTGAAGAAAAAGCAAAAAAGCAAGCGCAACAAGCCTCT 1236  
Db 1150 GGAGGAAGAAACATCGGTTGTGAAGAAAGCAAAAAAGCAAGCGCAACAAGCCTCT 1209  
Qy 1237 GCTGGATGAGATTTGACCTGTGTACCGCGGAGTGTCAACGAGGAAGTATATGACAGGAG 1296  
Db 1210 GCTGGATGAGATTTGCTGTGTACCGAGGAGTGTCTATGAGGAGGTTATGCTGGCAG 1269  
Qy 1297 CCACGAGTATCCAGGAGGAGTCTATCTCTCAAGTTTGATATTCCTACTCTCTCTGTG 1356  
Db 1270 CCATCAATATCCAGGAGGAGGTCATCTCTCAAGTTTGACAACTCTCTACTCTTTGTG 1329  
Qy 1357 GAGTCCCAAGTCCGCTACTACAGAGTCTATATACTAGATAGAGTCTGCTCTTCCAAGGT 1416  
Db 1330 CGGGTCAAAATCAGTCTACTACAGAGTCTATATACTAGATAAAATTTTGTGTACAAAGT 1389  
Qy 1417 CCGGAGTCCAGGGTTGACACAAACATGACGTTTAAATTT 1454  
Db 1390 CTGGAGTCTAGGGTTGGCAGAGATGACATTTAATTT 1427

RESULT 4

AAK53035  
ID AAK53035 standard; cDNA; 2350 BP.  
XX AC AAK53035;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human polynucleotide SEQ ID NO 2564.  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US04098.  
XX PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI: 2001-476283/51.  
DR P-PSDB; AAK79902.  
XX Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -  
XX Claim 1: Page 4841-4842; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX SQ Sequence 2350 BP; 687 A; 497 C; 628 G; 538 T; 0 other;

Query Match 71.7%; Score 1045.4; DB 22; Length 2350;  
Best Local Similarity 86.5%; Pred. No. 7.3e-271;  
Matches 1177; Conservative 0; Mismatches 181; Indels 3; Gaps 2;  
Qy 97 AGTAAAGATGGCAAGCCTTTTCATCCAACTTATGAAGAAAAACTGAAGTTGCTGGCACT 156  
Db 335 AGAAAAAGATGCCAAGCATTTTCATCCAACTTATGAAGAAAAATGAAGCTTGTGGCACT 394  
Qy 157 GCATAAGCAAGTTCTTTTGGGCCCATATAACCCAGACACGCTCCCTGAGTTGGATTCTT 216  
Db 395 GCATAAGCAAGTTCTTTTGGGCCCATATAATCCAGACACTTGTCTGAGTTGGATTCTT 454  
Qy 217 TGATGTGTTGGGAATGATAGGAGGAGAGAAATGGGACGCTCTGGCAACATCTCAAGA 276  
Db 455 TGATGTGTTGGGAATGATAGGAGGAGAGAAATGGGACGCTCTGGCAACATCTCTAAAGA 514  
Qy 277 GGATGCCATGTTAGAGTTTGTGAAGCTTCTAAATAGTGTCTCTCTCTCTCGGCATA 336  
Db 515 GGATGCCATGTTGGAGTTTGTCAAGCTCTTAATAGTGTGTTGCCATCTCTTTTCAACATA 574  
Qy 337 TGTTCGCTCCACAGATAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 396  
Db 575 TGTTCGCTCCACAGATAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634  
Qy 397 CGGAGGCGCTGAG 456  
Db 635 CGGAGGCGCTGAG 694  
Qy 457 GCGAGAGGAG 516  
Db 695 GAG 754  
Qy 517 GAGGCTTCGGTGGAG 576  
Db 755 AAGGCTTCGGTGGAG 814  
Qy 577 CGTCAATTCAGAGAGATGAG 636  
Db 815 CGTCAATTCAGAGAGATGAG 874  
Qy 637 TCTCATCCGCGCTGAG 696  
Db 875 TCTCATCCGCGCTGAG 934  
Qy 697 CCAACTCCACAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 756  
Db 935 CCAGCTTCAGAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 994  
Qy 757 ATCATTGCTGCATCATCAAGGTTGAACACAGCTGGAG--CAAGTGATACACTGTCTAG-T 813  
Db 995 TTCTTGTGCTCATCATCAAAAGTGAATGCAACTGTACCCCAAGTAATATGATGTCAAGTT 1054





Db 422 TTTTGGGAATTTGCCACAGACAAATATGACATGGGTTGGGTATATTTCAATGGACA 481  
QY 1121 GACTCTCCAAATGCTGCTGTCTAGTGTGATGTCAGTCCAGTGACGAGGAGGAG 1180  
Db 482 GACTCTCCAAACACTGCTGTACAGCTGCTGATGTCAGTGTGATGTCAGGATCCAGCAGGAG 541  
QY 1181 GAGGAAGAAATGTCATTTGTGAAGAAAAAGCAAAAAAGACGCCAACAGCCTTGCTG 1240  
Db 542 GAAGAAGAAACATCGGTGTGTGAAGAGAAAGCCAAAAAGATGCCAACAGCCTTTGCTG 601  
QY 1241 GATGAGATTGACCTGTGTACCGGGGACGTCTACGAGGAAGTATATGCAGCGAGCCAC 1300  
Db 602 GATGAGATTGCTGTGTACCGAGGACGTCTATGAGGAGGTATGCTGGCAGCCAT 661  
QY 1301 CAGTATCCAGGAGGGGAGTCTATCTCTCAAGTTTGATAATTCCTACTCTGTGGAGG 1360  
Db 662 CAATATCCAGGAGGAGGTCTATCTCTCAAGTTTGATAATTCCTACTCTGTGGCGG 721  
QY 1361 TCCAGTCCGCTACTACAGAGTCTATTATCTAGTATAGAGTGTCTTCCAGTCCGG 1420  
Db 722 TCAAAATCAGTCTACTACAGAGTCTATTATCTAGTAAAAATGTTGTACAAAGTCTGG 781  
QY 1421 AGTCCAGGTTGAGCACACATGACGTTTAATTT 1454  
Db 782 AGTCTAGGTTGGGAGAGATGACATTTAATTT 815

RESULT 6

AAC77363  
ID AAC77363 standard; cDNA; 420 BP.  
XX AC AAC77363;  
XX AC AAC77363;

08-FEB-2001 (first entry)

Human ORFX ORF2918 polynucleotide sequence SEQ ID NO:5835.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
hypotensive; dermatological; immunosuppressive; antiinflammatory;  
antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
cholesterol ester storage; systemic lupus erythematosus; infection;  
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
bone damage; cartilage damage; antiinflammatory disease; coagulation;  
thrombosis; contraceptive; ss.

Homo sapiens.

OS WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

DR P-PSDB; ABA43154.

XX Novel nucleic acids and peptides derived from open reading frame X.

PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 5; Page 5005; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX Sequence 420 BP; 134 A; 69 C; 127 G; 88 T; 2 other;

Query Match 20.1%; Score 293.2; DB 21; Length 420;  
Best Local Similarity 89.3%; Pred. No. 9.2e-69;  
Matches 316; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 97 AGTAAAGATGCGAAGCGCTTTCATCCAACTTATGAAGAAAACCTGAAGTTCGTGGCACT 156  
Db 66 AGAAAAGATGCGAAGCATTTTCATCCAACTTATGAAGAAAATTTGAAGCTTGTGGCACT 125  
QY 157 GCATGAAGCAAGTCTTTTGGGCCCATATAACCCAGACACGTCCTCCCTGAGTTGGATTCTT 216  
Db 126 GCATGAAGCAAGTCTTTTGGGCCCATATAATCCAGACACTTGTCTCTGAGTTGGATTCTT 185  
QY 217 TCATGCTGTGGGAATGATAGGAGGAGAGATGGCGAGCTCTGGGAACATGTCACAGA 276  
Db 186 TCATGCTGTGGGAATGATAGGAGGAGAGATGGCGAGCTCTGGGAACATGTCACAGA 245  
QY 277 GGATCCCATGGTAGAGTTTGTGAAGCTTCTAAATAGTGTCTCTCTCTCGGCATA 336  
Db 246 GGATCCCATGGTAGAGTTTGTGAAGCTTCTAAATAGTGTCTCTCTCTCGGCATA 305  
QY 337 TGTTCCTGCTCCACAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 396  
Db 306 TGTTCCTGCTCCACAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365  
QY 397 GCGAAGGCGCGTGAAG 450  
Db 366 GCGAAGGCGCGTGAAG 419

RESULT 7

AAK53762/c  
ID AAK53762 standard; cDNA; 278 BP.  
XX AC AAK53762;  
XX AC AAK53762;

16-NOV-2001 (first entry)

Murine transport and binding associated protein encoding cDNA SEQ ID 327.

Murine; liver; gene library; amino acid synthesis; binding protein;  
cell metabolism; energy metabolism; fatty acid metabolism; synthesis;  
phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;  
replication; transcription; translation; transport protein; ss.



[illegible]

QY 894 GACCAAAAG 902  
|||||  
Db 433 GACCAAAAG 441

RESULT 11  
ID AAK77371/c  
XX AAK77371 standard; DNA; 4384 BP.  
AC AAK77371;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32183.  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-MAR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225271.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 12-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.



```
XX 30-JAN-2001; 2001WO-US00669.
XX PF
XX PR
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 1; SEQ ID NO 11260; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 554 BP; 156 A; 149 C; 114 G; 135 T; 0 other;

Query Match 7.6%; Score 110.6; DB 22; Length 554;
Best Local Similarity 61.1%; Pred. No. 1.5e-19;
Matches 179; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 901 AGACTCTCTTCAGTGATTGCGAGCTCCATGCCATGTGGACAAGACCACAAATCAAAAGACTT 960
Db 483 AGTCTCCCACCTCTGATGGCTCTCCATGCATCTGGACCTTTGCCAAGGTGAAGAAATT 424

Qy 961 TAAAGAGAAGATTCCGCGAGGATTCAGATTCTGTGATTACAGTACGTCGAGGAGAAGTCGT 1020
Db 423 CAAAAGCAAGCTGGGCAAGAGAAGAACAGCGCTCTGGTGGTGAAGCGTGGTGAGGTGGT 364

Qy 1021 CACCGTCCGACTCCGACTCATGAGGAGGATCATACCTATTTTGGGAATTGGCCACAGA 1080
Db 363 GACCATCCGGGTACCTACTCATCCAGAGGGGAAGCGTGTCTGCTGGGAGTTGGCACCGA 304

Qy 1081 CAGTTATGACATTTGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTGT 1140
Db 303 TGACTATGACATTTGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTGT 244

Qy 1141 CAGTGTGATGTCAGTGCAGTCCAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1193
Db 243 AACTGTGCAGTGCAGTGCAGTCCAGTCCAGTCCAGTGCAGTGCAGTGCAGTGCAGTGCAG 191

RESULT 14
AAK11364/c
XX ID AAK11364 standard; DNA; 554 BP.
XX AC AAK11364;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 11355.
XX KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
```

```
OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PT analyzing gene expression in human
XX PS brains -
XX CC Example 4; SEQ ID NO: 11355; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX SQ Sequence 554 BP; 156 A; 149 C; 114 G; 135 T; 0 other;

Query Match 7.6%; Score 110.6; DB 22; Length 554;
Best Local Similarity 61.1%; Pred. No. 1.5e-19;
Matches 179; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 901 AGACTCTCTTCAGTGATTGCGAGCTCCATGCCATGTGGACAAGACCACAAATCAAAAGACTT 960
Db 483 AGTCTCCCACCTCTGATGGCTCTCCATGCATCTGGACCTTTGCCAAGGTGAAGAAATT 424

Qy 961 TAAAGAGAAGATTCCGCGAGGATTCAGATTCTGTGATTACAGTACGTCGAGGAGAAGTCGT 1020
Db 423 CAAAAGCAAGCTGGGCAAGAGAAGAACAGCGCTCTGGTGGTGAAGCGTGGTGAGGTGGT 364

Qy 1021 CACCGTCCGACTCCGACTCATGAGGAGGATCATACCTATTTTGGGAATTGGCCACAGA 1080
Db 363 GACCATCCGGGTACCTACTCATCCAGAGGGGAAGCGTGTCTGCTGGGAGTTGGCACCGA 304

Qy 1081 CAGTTATGACATTTGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTGT 1140
Db 303 TGACTATGACATTTGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTGT 244

Qy 1141 CAGTGTGATGTCAGTGCAGTCCAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1193
Db 243 AACTGTGCAGTGCAGTGCAGTCCAGTCCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 191

RESULT 15
AAK37154/c
XX ID AAK37154 standard; DNA; 554 BP.
XX AC AAK37154;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 11711.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
```

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

09-AUG-2001.  
PD  
XX

30-JAN-2001: 2001WO-US00668.

XX  
PR 04-FEB-2000: 2000US-0180312

PR 26-MAY-2000; 2000US-0207456.  
PR 30-MAY-2000; 2000US-0209408  
PR 30-MAY-2000; 2000US-0209408

PR 03-AUG-2000; 2000US-0632366.

PR 27-SEP-2000; 2000US-0236359.

[illegible]XX  
FA (MOLE-%) MOLECULAR DYNAMICS IPI Penn SG, Hanzel DK, Chen W  
XX

DR WPI; 2001-488900/53.  
vv

PT Human genome-derived single exon nuclei

XX  
 1947-1948

F3  
XX  
example 4; SEQ ID NO: 11/11; 030pp + sequence listing; English.

CC The present invention provides a number of single  
CC probes which are derived from genomic sequences

CC bone marrow. They can be used to measure gene expression in bone marrow samples which may enable the improved diagnosis and treatment of

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

XX  
SQ Sequence 554 BP; 156 A; 149 C; 114 G; 135 T; 0 other;

Query Match	7.6%	Score 110.6;	DB 22;	Length 554;
-------------	------	--------------	--------	-------------

Best Local Similarity 61.1%; Pred. No. 1.5e-19;  
Matches 179; Conservative 0; Mismatches 114; Indels 0; Gaps

QY 901 AGACTCTCTTCCAGTGATTCAGCTCCATCCATGTGGACAAGACCACAAATCAAGACTT 960

• Db 483 AGTCTCCCCACCTCTGATGGCTCCCTCCATGCATCTGGACCTTTGCCAAGGTGAAGGAATT 424

QY 961 TAACAGAGAAGATTCCGCAGGATGCACGATTCTGTGATTACAGTACGTCGAGGAGAGTCTCGT 1020

Dδ	423	C A A A G C A A G C T G G G C A A G A G A A C A G C C G T C T G T G T G A A G C T G T G A G G T G T	364
QY	1021	C A C G G T C C G A G T C C G A C T C A T G A G G A A G G A T C A T A C T A T T T T G G G A A T T T G C C A C A G A	1080
Dδ	363	G A C C A T C C G G G T A C C T A C T C A T C C A G A G G G A A G C G T G T C T G C T G G G A G T T T G G C A C C G A	304
QY	1081	C A G T T A T G A C A T T G G G T T T G G G G T T A T T T T G A A T G A C A C A C T C T C C A A A T G T G C T G T	1140
Dδ	303	T G A C T A T G A C A T T G G C T T T G G A G T T A T T T G A C T G A C C C C T G T A C T A G C A C T G A C A T	244
QY	1141	C A G T G T G C A T G T C A G T G A G T C C A G T A C G A G G A G G A G G A G A A A A T G	1193
Dδ	243	A A C T G T C A G G T C A G T G A T T C C A G T G A C G A T G A G G A T C A A G A A G A G A A G G	191

Search completed: February 21, 2003, 21:43:11  
Job time : 351 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 21:31:21 ; Search time 67 Seconds  
(without alignments)  
6678.230 Million cell updates/sec

Title: US-09-762-594-2  
Perfect score: 1459  
Sequence: 1 gaatcgccgcgcgtgcac.....catgaacttaatttccttt 1459

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	99.6	6.8	7218	1 US-08-232-463-14	Sequence 14, Appl
2	74.2	5.1	3489	2 US-08-728-323A-1	Sequence 1, Appl
3	74.2	5.1	3489	4 US-09-298-568-1	Sequence 1, Appl
C 4	74.2	5.1	32207	2 US-08-770-379-20	Sequence 20, Appl
C 5	74.2	5.1	32207	4 US-08-757-669A-20	Sequence 20, Appl
C 6	74.2	5.1	32207	4 US-09-230-371A-20	Sequence 20, Appl
7	72.6	5.0	1926	4 US-09-249-585A-2	Sequence 2, Appl
8	72.6	5.0	2580	3 US-09-050-863-2	Sequence 2, Appl
9	72.6	5.0	2580	4 US-09-359-081-2	Sequence 2, Appl
C 10	72.6	5.0	5452	2 US-09-130-114-1	Sequence 1, Appl
11	72.6	5.0	9600	4 US-08-910-647-1	Sequence 1, Appl
12	72.6	5.0	9600	4 US-09-620-925-1	Sequence 1, Appl
13	72.6	5.0	10596	1 US-07-884-811-15	Sequence 15, Appl
14	72.6	5.0	10596	1 US-07-885-971-15	Sequence 15, Appl
15	72.6	5.0	10596	1 US-08-087-783A-15	Sequence 15, Appl
16	72.6	5.0	10596	1 US-08-194-088B-15	Sequence 15, Appl
17	72.6	5.0	10596	2 US-08-194-087-15	Sequence 15, Appl
18	72.6	5.0	10596	5 PCT-US93-04648-15	Sequence 15, Appl
19	70.8	4.9	543	6 5273901-6	Patent No. 5273901
20	69.4	4.8	533	6 5482709-5	Patent No. 5482709
C 21	69.4	4.8	16442	3 US-08-781-891-208	Sequence 208, App
C 22	65.6	4.5	51259	3 US-08-781-891-209	Sequence 209, App
23	63.6	4.4	9551	1 US-08-056-200-93	Sequence 93, Appl
24	63.6	4.4	9551	2 US-08-800-644-93	Sequence 93, Appl
25	60.6	4.2	289	4 US-09-007-005-17	Sequence 17, Appl
26	60.6	4.2	289	4 US-09-244-796-17	Sequence 17, Appl
27	56.8	3.9	397	3 US-09-253-691-3	Sequence 3, Appl

28 56 3.8 4266 4 US-09-651-011A-3 Sequence 3, Appl  
29 54.8 3.8 939 1 US-08-285-440-10 Sequence 10, Appl  
30 54.8 3.8 939 1 US-08-630-349-10 Sequence 10, Appl  
31 54.8 3.8 1599 1 US-08-285-440-11 Sequence 11, Appl  
32 54.8 3.8 1599 1 US-08-630-349-11 Sequence 11, Appl  
33 54.8 3.8 1677 1 US-08-285-440-12 Sequence 12, Appl  
34 54.8 3.8 1677 1 US-08-630-349-12 Sequence 12, Appl  
35 54.8 3.8 2137 1 US-08-285-440-13 Sequence 13, Appl  
36 54.8 3.8 2137 1 US-08-630-349-13 Sequence 13, Appl  
37 54.8 3.8 2215 1 US-08-285-440-14 Sequence 14, Appl  
38 54.8 3.8 2215 1 US-08-630-349-14 Sequence 14, Appl  
39 54.4 3.7 248 4 US-09-007-005-32 Sequence 32, Appl  
40 54.4 3.7 248 4 US-09-244-796-32 Sequence 32, Appl  
41 54.4 3.7 277 4 US-09-007-005-3 Sequence 3, Appl  
42 54.4 3.7 277 4 US-09-244-796-3 Sequence 3, Appl  
43 54.4 3.7 2301 1 US-08-306-691B-23 Sequence 23, Appl  
44 54.4 3.7 2301 5 PCT-US93-06251-78 Sequence 78, Appl  
45 53.8 3.7 152331 3 US-09-128-155-16 Sequence 16, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 6.8%; Score 99.6; DB 1; Length 7218;





[illegible]





```
QY 656 GAGCAGCACTATCAGCACTATAACACCAGGAGAGCAAAACCCAACTTCACACACAG 715
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1238 GAGCAGGAGGAGGGGCGAGGAGGGCGAGGAGGAGGAGGGCGAGGAGGCGAGGAG 1297
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 716 GCAGCATTTACAGAAACAGCAAGAGTACTGATGCTGGG 754
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1298 GAGGGCGAGGAGGCGAGGAGCAGGAGGCGAGGAGGGG 1336
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/620.925
; FILING DATE: 21-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910.647
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38.459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

Query Match 5.0%; Score 72.6; DB 4; Length 9600;
Best Local Similarity 48.9%; Pred. No. 2.8e-10;
Matches 195; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 356 GAGAGGAGAGAGAGAGCAAAAGAGAGGCGGAGGAGGAGCGGCGGTGAAGAG 415
||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 938 GAGCAGGAGGAGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGAGGAGGAGGGG 997
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 416 GAAGAACGAGAGCGCTGCAAAAGGAAGAGCGGAAGCGGAGGAGAGGAGCGG 475
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 998 CAGGAGGGCGAGGAGGGCGAGGAGCAGGAGGAGGAGGGCGAGGAGGCGAGGAGGGG 1057
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 476 CTGAGACGGGAGGAGAGAGAGCGCGGATAGAGAGAGAGAGGCTTCGGCTGGAACAG 535
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1058 CAGGAGGGCGAGGAGGAGGCGAGGAGCAGGAGGAGGAGGAGGCGAGGAGGGG 1117
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 536 CAAAAGCAGCAGATATTCGAGCTTTAACTCGCAGACTGCCGTGCGATTCACAGATAT 595
|| | ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 1118 CAGGAGCAGGAGGGCGAGGAGCAGGAGGGGCGAGGAGGAGGAGGAGGAGGGG 1177
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 596 GCAGCCCGCAGCATATCCAGGGAACACTACAAACACAGCAGAGATTCTATCCCGCAGTGCAG 655
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1178 CAGCAGGGCGAGGAGCAGGAGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGG 1237
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 656 GAGCAGCACTATCAGCACTATAACACCAGGAGAGCAAAACCCAACTTCACACACAG 715
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1238 GAGCAGGAGGAGGGCGAGGAGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1297
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 716 GCAGCATTTACAGAAACAGCAAGAGTACTGATGCTGGG 754
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1298 GAGGGCGAGGAGGCGAGGAGCAGGAGGCGAGGAGGGG 1336
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA: US/07/884.811
; APPLICATION NUMBER: US/07/884.811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-884-811-15
```

```
Query Match 5.0%; Score 72.6; DB 1; Length 10596;
Best Local Similarity 48.9%; Pred. No. 3e-10;
Matches 195; Conservative 0; Mismatches 204; Indels 0; Gaps 0;
```

```
QY 356 GAGAGGAGAGAGAGAGCAAAAGAGGCGGAGGAGGAGCGGAGGAGCGGTGAAGAG 415
||| |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 2472 GAGCAGGAGGAGGGCGAGGAGCAGGAGGAGGAGGGCGAGGAGGAGGAGGAGGAGGGG 2531
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 416 GAAGAACGAGAGCGGCTGCAAAAGGAAGAGAGCGGAAGCGGAGGAGAGAGACCGG 475
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2532 CAGGAGGGCGAGGAGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGAGGAGGAGGGG 2591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 476 CTGAGACGGGAGGAGAGAGAGCGCGGATAGAGGAGAGAGAGGCTTCGGCTGGAACAG 535
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2592 CAGGAGGGCGAGGAGCAGGAGGAGGGCGAGGAGGAGGAGGAGGAGGAGGGG 2651
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```







**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	204.4	14.0	934	10	US-09-866-562-33		Sequence 33, Appl
2	113.4	7.8	527	10	US-09-866-562-32		Sequence 32, Appl
3	110.6	7.6	554	10	US-09-864-761-13381		Sequence 13381, A
C 4	109.2	7.5	306	10	US-09-864-761-30285		Sequence 30285, A
5	69.4	4.8	720	9	US-10-101-487-74		Sequence 74, Appl
c 6	69.4	4.8	720	9	US-10-101-487-76		Sequence 76, Appl
7	66.8	4.6	575	10	US-09-864-761-20733		Sequence 20733, A
8	66.8	4.6	1969	9	US-09-864-761-3972		Sequence 3972, Ap
c 9	65	4.5	659158	9	US-09-771-208-20		Sequence 20, Appl
10	63.4	4.3	3809	12	US-10-001-870-68		Sequence 68, Appl
c 11	62.8	4.3	2108	10	US-09-862-832-225		Sequence 225, App
C 12	62.4	4.3	474	10	US-09-864-761-11284		Sequence 11284, A
13	62	4.2	522	9	US-10-101-487-71		Sequence 71, Appl
C 14	62	4.2	530	9	US-10-101-487-73		Sequence 73, Appl
15	62	4.2	554	9	US-10-101-487-69		Sequence 69, Appl
16	62	4.2	554	9	US-10-101-487-106		Sequence 106, App
C 17	61.8	4.2	475	10	US-09-864-761-1361		Sequence 1361, Ap
C 18	61.8	4.2	512	10	US-09-864-761-18121		Sequence 18121, A
c 19	61	4.2	475	10	US-09-864-761-6203		Sequence 6203, Ap



```

: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Aomics-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 30285
: LENGTH: 306
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO ACO07954.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
: OTHER INFORMATION: NT HIT: AF022770.2, EVALUOE 3.00e-03
: OTHER INFORMATION: SWISSPROT HIT: P49193, EVALUOE 1.30e-01
: OTHER INFORMATION: EST_HUMAN HIT: AW292776.1, EVALUOE 0.00e+00
: US-09-864-761-30285

```

	Query Match	7.5%	Score 109.2	DB 10	Length 306
	Best Local Similarity	61.0%	Pred. No. 2.1e-20		
	Matches 177	Conservative	0	Mismatches 113	Indels 0
				Gaps 0	
QY	904	CTCTCTTCAGTGGATTCGAGCTCCATCCATGTGCAGACAGACACACAAATCAAGACTTTAA	963		
Db	305	CTCCCCACCTCTGTATGGTCTCTCATGTCATCTGCACCTTTGCCAAGGTGAAGGAATTCAA	246		
QY	964	AGAGAAGATTCCGCCAGATGACAGATTCTGTGATTACAGTACGTCGCGAGAGAGTCGTCCAC	1023		

D	b		245	AAGCAAGCTGGGCAAGAAGAACACGCCGTCTGGTGGTTGAAGCGGTGGTGAGGTGGTGAC	186
Q	y		1024	CGTCGGAGTCCCGACTCATATGAGGAAGGATCATACCTATTATTTGGGAATTTGGCACACACAG	1083
D	b		185	CATCGGGTACTCTACTCACCAGAGGGGAAGCGTGTCTCTGGGAGTTTGGCACCAGTGA	126
Q	y		1084	TTATGACAATGGGTTTTGGGGTTTATTTTAACTGCACAGACTCTCCAATTCCTGCTCTCAG	1143
D	b		125	CTATGACATGCTTTGGAGTTTATTTGACTGGACCCCTGTAAC TAGCACTGACATAAC	66
Q	y		1144	TGTGATGTCAGTGAGTCCAGTCGACAGGAGGAGGAGGAGGAGAAAATG	1193
D	b		65	TCTGCAGGTCACTGATTCAGTCAGCATGAGGATGAAGAAGAGGAGG	16
RESULT 5					
US-10-101-487-74					
; Sequence 74, Application US/10101487					
; Patent No. US20020169125A1					
; GENERAL INFORMATION:					
; APPLICANT: LEUNG, DAVID W.					
; APPLICANT: BERGMAN, PHILIP A.					
; APPLICANT: LOFUQUST, ALAN					
; APPLICANT: PIETZ, GREGORY E.					
; APPLICANT: TOMPKINS, CHRISTOPHER K.					
; APPLICANT: WAGONER JR., DAVID W.					
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES					
; TITLE OF INVENTION: THEREOF					
; FILE REFERENCE: 077319/0329					
; CURRENT APPLICATION NUMBER: US/101,487					
; CURRENT FILING DATE: 2002-03-20					
; PRIOR APPLICATION NUMBER: 60/277,705					
; PRIOR FILING DATE: 2001-03-21					
; NUMBER OF SEQ ID NOS: 116					
; SOFTWARE: PatentIn Ver. 2.1					
; SEQ ID NO 74					
; LENGTH: 720					
; TYPE: DNA					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic					
; OTHER INFORMATION: oligonucleotide					
; FEATURE:					
; NAME/KEY: CDS					
; LOCATION: (2)..(718)					
; US-10-101-487-74					

	Query Match	4.8%;	Score 69.4;	DB 9;	Length 720;	
	Best Local Similarity	48.6%;	Pred. No. 3e-09;			
	Matches 190;	Conservative 0;	Mismatches 201;	Indels 0;	Gaps 0;	
Qy	355	AGAGNAGGAAGAAGAAAGAAAGAACGCCGAGGAGGACCAAGCGCAGCTGAACA	414			
Dd	319	AGAGGAGGAAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGA	378			
Qy	415	GGAGAACCAGGCGCCTCAAAAGGAAGAAGAAAGCGGAAGCCAGAGGAGGAAGACCG	474			
Dd	379	GGAGAGGAGGAAGGAGGAGGAAGGAAGAGGAGGAAGGAGGAGGAAGAGGAGGA	438			
Qy	475	GCTGAGACGGGAGGAGGAGGAGCGCGGTATAGAGGAAGAGAGGCTTTCGGCTGGAACA	534			
Dd	439	AGAGSAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	498			
Qy	535	GCAAAGGACAGCATATATGGCAGCTTTTAACCTGCAGACTGCCGTGCAATTCCACAGTA	594			
Dd	499	GGAAGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA	558			
Qy	595	TGACGCCAGCAGTATCCAGGGAACTACGACAACAGCAGGATTTCTATCCCGCACCTGCA	654			
Dd	559	AGAGGAGGAAGGAAGAGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGA	618			
Qy	655	GGAGCAGCACTATCAGCAGTATAACACCCAGCAGCAAGAACCCCAACCTTGCAACAACA	714			











Db 67 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 126  
Qy 416 GAAGAACGAGAGCGGCTGCAAAAGGAAGAAGAAGAGCGGGAAGCGGAGAGGAAGACCGG 475  
Db 127 GAAGAGGAAGAGAGGAGGAAGAAGAGAGGAGGAGGAGGAGGAAGAAGAG 186  
Qy 476 CTGAGACGGAGGAGGAAGAGAGGGGGGATAGAGGAAGAGAGCTTCGGGTGGAACAG 535  
Db 187 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 246  
Qy 536 CAAAAGCAGCAGATATGCCAGCTTTAACTCCGAGACTCGCGTCAATTCCAGCAGTAT 595  
Db 247 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 306  
Qy 596 GCAGCCAGCAGTATCCAGGGAACTACGAACAACAGCAGATTTCTCATCGCCAGCTGCAG 655  
Db 307 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAAGAAGAG 366  
Qy 656 GAGCAGCACTATCAGCAGTATTAACCAACAGGCGAGCAACCAACCTGCACAACAACAG 715  
Db 367 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 426  
Qy 716 GCAGCATTTACAAACAGCAAGAAGTAGTG 745  
Db 427 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAG 456

RESULT 14  
US-10-101-487-73/c  
; Sequence 73, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOQUIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR FILING DATE: 60/277,705  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 73  
; LENGTH: 530  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: oligonucleotide  
US-10-101-487-73

Query Match 4.2%; Score 62; DB 9; Length 530;  
Best Local Similarity 47.4%; Pred. No. 2.8e-07;  
Matches 185; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

Qy 356 GAGAAGGAAGAAGAGAAAAAGAGCGGAGGAGGAGCGAAGGCGGTGAAGAG 415  
Db 464 GAAGAGGAAGAGAGGAGGAAGAAGAAGAGGAGGAGGAGGAGGAAGAAGAG 405  
Qy 416 GAAGAACGAGAGCGGCTGCAAAAGGAAGAAGAGCGGAGCGGAGAGGAAGACCGG 475  
Db 404 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 345  
Qy 476 CTGAGACGGAGGAGGAAGAGAGGGCGGATAGAGGAAGAGAGGCTTCGGGTGGAACAG 535  
Db 344 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 285  
Qy 536 CAAAAGCAGCAGATATGCCAGCTTTAACTCCGAGACTCGCGTCAATTCCAGCAGTAT 595

Db 284 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 225  
Qy 596 GCAGCCAGCAGTATCCAGGGAACTACGAACAACAGCAGATTTCTCATCGCCAGCTGCAG 655  
Db 224 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 165  
Qy 656 GAGCAGCACTATCAGCAGTATATAACACCAGGCGAGAGCAAAACCCACCTGCACAACAACAG 715  
Db 164 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 105  
Qy 716 GCAGCATTTACAGAAACAGCAAGAAGTAGTG 745  
Db 104 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAG 75

RESULT 15  
US-10-101-487-69  
; Sequence 69, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOQUIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR FILING DATE: 60/277,705  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 69  
; LENGTH: 554  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: oligonucleotide  
; NAME/KEY: CDS  
; LOCATION: (1)..(528)  
US-10-101-487-69

Query Match 4.2%; Score 62; DB 9; Length 554;  
Best Local Similarity 47.0%; Pred. No. 2.9e-07;  
Matches 191; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

Qy 356 GAGAAGGAAGAAGAGAAAAAGAGCGGAGGAGGAGCGAAGGCGGTGAAGAG 415  
Db 139 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 198  
Qy 416 GAAGAACGAGAGCGCTCCAAAAGGAAGAAGAGCGGAGCGGAGAGCGAAGCAACCGG 475  
Db 199 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 258  
Qy 476 CTGAGACGGAGGAGGAAGAGAGGGCGGATAGAGGAAGAGAGGCTTCGGCTGGAACAG 535  
Db 259 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 318  
Qy 536 CAAAAGCAGCAGATATGCCAGCTTTAACTCGCAGACTGCGGTGCAATTCCAGCAGTAT 595  
Db 319 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 378  
Qy 596 GCAGCCAGCAGTATCCAGGGAACTACGAACAACAGCAGATTTCTCATCGCCAGCTGCAG 655  
Db 379 GAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAGGAAGAAGAGGAGGAAGAAGAG 438  
Qy 656 GAGCAGCACTATCAGCAGTATATAACACCAGGCGAGCAAAACCCACCTGCACAACAACAG 715

```

Db 439 GAAGAGGAAGAGAGGAGGAGGAAGAGAGGAGGAGGAAGAGAGGAGGAGGAAGAGAGGAG 498
Qy 716 GCAGCATTACAGAAACAGCAAGAGTAGTGATGGCTGGGGGCATCAT 761
Db 499 GAAGAGGAAGAGAGGAGGAGGAAGAGAGGAGGAGGAGGAGTAGTCTT 544
```

Search completed: February 21, 2003, 23:59:11  
Job time : 1928 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 21, 2003, 21:36:46 ; Search time 2217 Seconds  
(without alignments)  
10658.194 Million cell updates/sec

Title: us-09-762-594-2  
Perfect score: 1459  
Sequence: 1 gaattcgccgcgcgcgcac.....catgacgttaatttccttt 1459

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	597.4	40.9 757 12	BG294067 602390930
2	555.6	38.1 932 14	BQ223172 AGENCOURT
3	542.4	37.2 955 9	AL546667 AL546667
4	499.8	34.3 1108 12	BG254119 602367007
5	491.2	33.7 737 12	BG776473 602663560
6	487.6	33.4 707 12	BG705952 602669271

7	485.6	33.3	493	10	BB704602	BB704602
8	452.4	31.0	600	13	BI990454	BI990454
9	452.2	31.0	625	10	AW963720	AW963720
10	445.4	30.5	649	12	BE916087	BE916087
11	431.4	29.6	984	12	BF797451	BF797451
12	405	27.8	519	12	BE694778	BE694778
13	390.4	26.8	682	14	BQ180694	BQ180694
14	368.6	25.3	1114	14	BM804990	BM804990
15	367.4	25.2	639	12	BG533826	BG533826
16	366.4	25.1	489	10	AW631077	AW631077
17	361.4	24.8	363	12	BF226428	BF226428
18	350.8	24.0	449	14	W00483	W00483
19	349.8	24.0	461	13	BI339401	BI339401
20	346.4	23.7	774	9	AL529297	AL529297
21	343.8	23.6	395	9	AT840648	AT840648
22	339.4	23.3	592	14	BQ781415	BQ781415
23	327.6	22.5	774	10	BE541801	BE541801
24	326.2	22.4	895	14	BM801531	BM801531
25	316	21.7	332	10	BB706613	BB706613
26	311.4	21.3	476	10	BE142877	BE142877
27	307.4	21.1	434	9	AT604009	AT604009
28	301	20.6	462	14	N31282	N31282
29	294	20.2	445	14	T86543	T86543
30	293.2	20.1	614	9	AL679623	AL679623
31	292.4	20.0	679	13	BJ530032	BJ530032
32	289.2	19.8	842	12	BF786885	BF786885
33	284	19.5	390	10	BE185547	BE185547
34	282.8	19.4	729	12	BF035473	BF035473
35	282	19.3	421	9	AA636630	AA636630
36	278.6	19.1	431	12	BG080930	BG080930
37	277	19.0	828	13	BI868335	BI868335
38	273.6	18.8	692	10	BB617641	BB617641
39	273.2	18.7	450	9	AT506318	AT506318
40	271.6	18.6	388	9	AA148726	AA148726
41	271.4	18.6	496	13	BI965035	BI965035
42	267.4	18.3	390	14	C87779	C87779
43	267.4	18.3	407	14	C85460	C85460
44	267.4	18.3	411	14	C85479	C85479
45	267.4	18.3	415	12	BG067872	BG067872

ALIGNMENTS

RESULT 1  
BG294067  
LOCUS 602390930f1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:4502911 5', linear EST 21-FEB-2001  
DEFINITION mRNA sequence.  
ACCESSION BG294067.1 GI:13054331  
VERSION BG294067  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 757)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10372 row: k column: 08  
High quality sequence stop: 680.  
Location/Qualifiers 1. .757  
FEATURES source

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4502911"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      220 a   160 c   195 g   182 t
ORIGIN
Query Match      40.9%; Score 597.4; DB 12; Length 757;
Best Local Similarity 96.6%; Pred. No. 5.9e-131;
Matches 679; Conservative 0; Mismatches 6; Indels 18; Gaps 6;
QY 758 TCATTGCGCTGCATCATCAAAAGGTGAACACAGCTGGAGCAAGTGATACACTGTCACTTAAT 817
Db 1 TCATTGCGCTGCATCATCAAAAGGTGAACACAGCTGGAGCAAGTGATACACTGTCACTTAAT 60
QY 818 GCACAGGCCAARACCCACACTCAAAATCCGAAAAAGTCCCTTGAGCCAGAGCTCCAGAA 877
Db 61 GGACAGGGCCAAACCCACACTGAAAATCCGAAAAAGTCCCTTGAGCCAGAGCTCCAGAA 120
QY 878 GAAGCCTTTGGAAATGGACCAACAGACTCTCTCCAGTGAATGCAGCTCCATCCATGTTGG 937
Db 121 GAAGCCTTTGGAAATGGACCAACAGACTCTCTCCAGTGAATGCAGCTCCATCCATGTTGG 180
QY 938 ACAAGACCACAAATCAAGACTTTAAAGAGAGATTCGGCAGGATGCAGATCTCTGTGATT 997
Db 181 ACAAGACCACAAATCAAGACTTTAAAGAGAGATTCGGCAGGATGCAGATCTCTGTGATT 240
QY 998 ACAGTACGTCGAGGAGAAAGTCTCCACCGTCCGAGTCCGAGTCCATGAGAGGATCATAC 1057
Db 241 ACAGTACGTCGAGGAGAAAGTCTCCACCGTCCGAGTCCGAGTCCATGAGAGGATCATAC 300
QY 1058 CTATTTTGGGAATTTGCCACACAGACTTATGACATGGGTGTTGGGTATTTTGAATGG 1117
Db 301 CTA-TTTGGGAATTTGCCACACAGACTTATGACAT--GGTGGGTATTTTGAATGG 357
QY 1118 ACAGACTCTCCAAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1177
Db 358 ACAGACTCTCCAAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 405
QY 1178 GAGGAGGAGAAATGTCACCTTGTAAGAAAAAGCAAAAGAGAGCCCAACAGCCTCTG 1237
Db 406 GAGGAGGAGAAATGTCACCTTGTAAGAAAAAGCAAAAGAGAGCCCAACAGCCTCTG 465
QY 1238 CTGGATGAGATTGTACCTGTGTACCGCGGGACTGTACAGGAGAGTATATGCGAGGCAGC 1297
Db 466 CTGGATGAGATTGTACCTGTGTACCGCGGGACTGTACAGGAGAGTATATGCGAGGCAGC 525
QY 1298 CACAGATATCCAGGAGGGAGTCTATCTCTCAAGTTTGATTAATCTCTCTCTCTGTTGG 1357
Db 526 CACAGATATCCA-GGAGGGAGTCTATCTCTCAAG-TTGATAATCTCTCTCTCTCTGTTGG 583
QY 1358 AGGTCCAAAGTCGCTCTACTACAGAGCTTATATACAGAGCTGCTGTCTTCCAGGCTC 1417
Db 584 AGGTCCAAAGTCGCTCTACTACAGAGCTTATATACAGAGCTGCTGTCTTCCAGGCTC 643
QY 1418 CGGAGTCCAGGG-TTGAGCAACAACATGACGCTTTAATTTTCCCTTT 1459
Db 644 GGGAGTCTAGGGTTTGAGCACACATGACGCTTTAATTTTCCCTTT 686
RESULT 2
LOCUS BQ223172
DEFINITION AGENCOURT_7551342 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6066023
5', mRNA sequence.
ACCESSION BQ223172
```

```
VERSION BQ223172.1 GI:20404572
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 932)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI3343 row: d column: 24
High quality sequence start: 20
High quality sequence stop: 395.
FEATURES
source
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6066023"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 297 a 197 c 264 g 174 t
ORIGIN
Query Match 38.1%; Score 555.6; DB 14; Length 932;
Best Local Similarity 82.8%; Pred. No. 4.7e-121;
Matches 671; Conservative 0; Mismatches 134; Indels 5; Gaps 3;
QY 97 AGTAAAGATGCGAAGCTTTTCATCCAACTTATGAAGAAAACATGAGTTCGTGGCACT 156
Db 116 AGAAAAGATGCGAAGCATTTTCATCCAACTTATGAAGAAAACATTTGAGCTTTGTGGCACT 175
QY 157 GCATAAGCAAGTCTTTTGGGCCATATAACCCAGACAGTCCCTGAGGTGGATTCTT 216
Db 176 GCATAAGCAAGTCTTTTGGGCCATATAATCCAGACACTTGTCTGAGGTGGATTCTT 235
QY 217 TCATGCTGTGGGAATGATAGGAGAGAGATGGGACAGCTCTGGGAAACATGTCCAAGGA 276
Db 236 TGATGCTGTGGGAATGACAGGAGAGAGATGGGACCCCTGGGAAACATGTCTAAGA 295
QY 277 GGATGCCATGTAGAGTTGTGAAGCTTCTAAATAAGTGTGTCTCTCTCTCTCGGCATA 336
Db 296 GGATGCCATGTGGAGTTGTGAAGCTTCTAAATAAGTGTGTGTCTCTCTCTCTCAACATA 355
QY 337 TGTTCGCTCCACAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 396
Db 356 TGTTCGCTCCACAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 415
QY 397 GCGAAGGCGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 456
Db 416 GCGAAGGCGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
QY 457 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 516
Db 476 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
QY 517 GAGGCTTCGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 576
Db 536 AAGGCTTCGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 595
```

```
QY 577 CQTGCAATTCAGCAGTATGAGCCCGCAGATATCCAGGGAACTACGAACACAGCAGAT 636
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 596 CQTGAGTTTCGGCAGTATGGGCGCCCAACTATCCAGGAACTACGAACACAGCAGCAAT 655
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 637 TCTCATCCGCGAGTCGAGGAGCAGCAGTATCAGCAGTATTAACACACAGCAGCAGCAAC 696
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 656 TCTCATCCGCGAGTTGCGGAGCAACATATCAGCAGTATCAGCAGTATGATCAAGT 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 697 CCAACCTGCACAACACAGGAGCAGTATCAGAAACAGCAAGAGTAGTGGCTGGGGC 756
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 716 CCAGCTTTCACGGCAACAGCAGCAGTATCCGGAACACCGGAAGTAGTAGTGGCTGG 775
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 ATCATTTGCTGATCATCAAGGTGAACACAGCTGG-AGCAAGTAGTACATGTCTAG-TT 814
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 776 TTCTTTGGCTACTCATCAAAAGTGAATGCAACTGGTACCAGGTAAATGATGATGCTCGGTTT 835
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 815 AATGACAGAGCCCAAAACCCACACTGAAATTCGGAAGAGTCTTGTAGCC---AGAAGCT 871
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 836 AAGGACAGAGCCCAAAACACACTGACCGTTCCGAAAAAGAACTGGAACACGAAAGCTT 895
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 872 GCAGAAGAGAGCTTTGAAAAATGGACCAAAA 901
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 896 GCAAAAAACCCTCGGAAATGGGACCAA 925
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
AL546667 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1029YJ06 5
LOCUS prime.mRNA sequence.
DEFINITION AL546667 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1029YJ06 5
prime.mRNA sequence.
ACCESSION AL546667
VERSION AL546667.1 GI:12880005
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 955)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source 1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1029YJ06"
/tissue_type="placenta"
/clone_lib="LTI_NFL006_PL2"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 268 a 221 c 305 g 159 t 2 others
ORIGIN

Query Match 37.2%; Score 542.4; DB 9; Length 955;
Best Local Similarity 87.0%; Pred. No. 6.4e-118;
Matches 594; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

QY 97 AGTAAAGATGCAAGAGCTTTTCATCCAACTATCAAGAAAACTGAAGTTCTGGGCACT 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 AGAAAGATGCAAGAGCTTTTCATCCAACTATCAAGAAAAATGAAGCTTTGGGCACT 332
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 157 GCATAAGCAAGTCTTTTGGGCCCATATTAACCCAGACACAGCTCCCTGAGGTGGATTCTT 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 GCATAAGCAAGTCTTTTGGGCCCATATTAATCCAGACACTTGTCTGAGGTGGATTCTT 392
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 TGATGTGTTGGGAATGATAGGAGGAGAAATGGCAGCTCTGGGAACATGTCTCAAGAT 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 TGATGTGTTGGGAATGACAGGAGGAGAAATGGCAGCTCTGGGAACATGTCTCAAGAT 452
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 GGATGCCATGTTAGAGTTTGTGAAGCTTCTAAATAAGTGTGTCTCTCTCTCTCCGCGATA 336
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 453 GGATGCCATGTTGAGTGTGTCAGCTCTTAATATAGTGTGTGTCATCTCTCTTTCACATA 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 TGTTCGCTCCACAGATATAGAGAAGGAAGAAAGAGAGAAAAAGAGCGGAGGAGGA 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 513 TGTTCGCTCCACAAAAATAGAGAAGGAAGAGACAAAAAGGAGGAGGAAGAGGA 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 397 GCGAAGGAGCGGTCAAGAGGAGAACAGAGAGCGGCTCCAAAAAGAAAGAGAGCGGAA 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 GCGAAGGCGGCTGAAGAGGAGAAAGAGAACGCTCTGCAAAAGGAGGAGGAAGACGTAG 632
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 457 GCGAGAGGAGGAAGACCGGCTGAGACGGGAGGAGAGAGCGGCGGATAGAGGAAGA 516
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 633 GAGAGAAGAGAGGAAGAGGCTTCGACGGGAGAGAGAGGAAAGGAGCGGATAGAGAAGA 692
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 517 GAGCTTGGCTGGAACAGCAAGCAAGAGCAGATATATGCGAGCTTTAAACTCCGAGACTGC 576
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 693 AAGCTTCGGTTGGAGCAGCAAAAAGCAGCAGATAATGCGAGCTTTAAACTCCGAGACTGC 752
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 577 CGTCAATTCAGCAGTATGAGCGCCAGCAGTATCCAGGGAACTACGNAACAGCAGAGAT 636
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 753 CGTCAATTCAGCAGTATGAGCGCCCAACAGTATCCAGGGAACTACGNAACAGCAGCAAT 812
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 637 TCTCATCGCCAGCTGCGAGGAGCAGCAGTATCAGCAGTATATAACACAGGAGGAGCAAC 696
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 813 TCTCATCGCCAGCTGCGAGGAGCAACACTATCAGCAGTATCAGCAGCTGTATCAAGT 872
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 697 CCAACCTGCACAACACAGGAGCAGCATTTACAGAAACAGCAAGAGTAGTGGTGGGGC 756
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 873 CCAGCTGCACAGCAACAGGAGCAGCATTTACAGAAACACAGCAAGTAGTAGTGGTGGGTC 932
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 ATCATTTGCTGCATCACTCAAAAGG 779
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 933 TTCTTCGCTACATCATCAAAAG 955
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
BG254119 1108 bp mRNA linear EST 13-FEB-2001
LOCUS 602367007F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4475329 5',
DEFINITION mRNA sequence.
ACCESSION BG254119
VERSION BG254119.1 GI:12763935
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1108)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10300 row: n column: 02
High quality sequence stop: 640.
Location/Qualifiers

FEATURES
```

```
source
1..1108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4475329"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 465 a 199 c 310 g 133 t 1 others
ORIGIN
Query Match 34.3%; Score 499.8; DB 12; Length 1108;
Best Local Similarity 84.8%; Pred. No. 7.9e-108;
Matches 584; Conservative 0; Mismatches 102; Indels 3; Gaps 2;
QY 388 GGAGGAGGAGCGGCGGTGAAGAGGAGGAGGAGGCGGCTGCACAAAGGAAAGAAGA 447
DB 1 GGAAGAGGAGCGGCGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
QY 448 GAAGCGGAAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 507
DB 61 GAAAGCGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 508 ACAGGAAGAGGCGTCCGCTGGAACACACAAAGCAGCAGATATAGGCGAGCTTTAAACTC 567
DB 121 AGAAGAAGAAAGGCTTCGGTTGGAGCAGCAAAAGCAGCAGATATAGGCGAGCTTTAAACTC 180
QY 568 GCAGACTGCGGTGCATATCCAGC-AGTATGCAGCCGAGCAGTATCCAGGAACTACGAAC 626
DB 181 CCAGACTGCGGTGCATATCCAGC-AGTATGCAGCCGAGCAGTATCCAGGAACTACGAAC 240
QY 627 AACAGAGATTCATATCCGCGAGCTGCAGGAGCAGCATATCAGCAGTATATAACACCAAGG 686
DB 241 AGCAGCAAAATTCATATCCGCGAGCTGCAGGAGCAGCATATCAGCAGTATACATGCAAGCAGT 300
QY 687 CAGAGCAACCAACCTGCACACACAGGAGCAGTATACAGAAACAGCAAGAAGTAGTGA 746
DB 301 TGATCAAGTCCAGCTTGACAGCAGCAGGAGCAGTATACAGAAACAGCAAGAAGTAGTGA 360
QY 747 TGGCTGGGCGATCATATGCTGCATCATCAAAAGGTGAACACAGCTGGAGCAAGTGATACAC 806
DB 361 TGGCTGGGCTTCCTTCCTATCATCATCAAAAGTGAATGCAACTGTACCAAGTAATATGA 420
QY 807 TGTCAAGTAAATGGAGCAGGCAAAACCCACACTGAAAAATTCGAAAAAGTCTCTTGAGCCAG 866
DB 421 TGTCAAGTAAATGGAGCAGGCAAAACACACTGACAGCTCCGAAAAAGAACTGGAACCAAG 480
QY 867 AAGCTGCAGAAAGCGCTTGAAAAATGGACCAAAAGACTCTCTCCAGTGTATGCGAGTGC 926
DB 481 AAGCTGCAGAAAGCGCTTGAAAAATGGACCAAAAGACTCTCTCCAGTGTATGCGAGTGC 540
QY 927 CATCATGTGGACAAAGACCAAAATCAAAAGACTTTTAAAGAGAAGATTCGCGAGGATGCAG 986
DB 541 CATCATGTGGACAGCACTTCAGATCAAAAGACTTCAAAAGAGAAGATTCAGCAGGATGCAG 600
QY 987 ATTCTGTGATACAGTACGTGAGGAGAGGCTGCACCGTCCGA---GTCCGCACTATGA 1044
DB 601 ATTCAAGTATACAGTGGCGGAGGAGAGGAGTGGTCACTGCTACGAAAGGTAAACCAACCAATGA 660
QY 1045 GGAAGGATCATACCTTATTTGGGAATTGG 1073
DB 661 AGAAGGATCATATCTCTAAGGGGAAATTGG 689
RESULT 5
BG776473 737 bp mRNA linear EST 15-MAY-2001
LOCUS 602663560F1 NIH_MGC_59 Homo sapiens cdna clone IMAGE:4808871 5',
DEFINITION mRNA sequence.
```

```
ACCESSION BG776473
VERSION BG776473.1
KEYWORDS GI:14046777
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI662 row: g column: 16
High quality sequence stop: 704.
FEATURES
location/Qualifiers
1..737
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4808871"
/clone_lib="NIH_MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgccc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 222 a 140 c 186 g 189 t
ORIGIN
Query Match 33.7%; Score 491.2; DB 12; Length 737;
Best Local Similarity 87.3%; Pred. No. 8.5e-106;
Matches 538; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 839 GAAAAATTCGAAAAAGCTCTTTCAGCCAGAGCTGCAGAAAGAGCTTGGAAAAATGGACCA 898
DB 1 GACAGCTCCGAAAAAGAACTGGAACCAAGAGCTGCAGAAAGAGCTTGGAGAAATGGACCA 60
QY 899 AAAGACTCTCTTCAGTGATTTGCGAGTCCATCCATGTGGCAAGACCAACAAATCAAGAC 958
DB 61 AAAGAAATCTCTTCAGTAAATAGCAGCTCCATCCATGTGGACACGACCTCAGATCAAGAC 120
QY 959 TTAAAGAGAGAGATTCGCGCAGGATCCAGATTCGTGATTACAGTACAGTCGAGGAGAGCTC 1018
DB 121 TTCAAGAGAGAGATTCAGCAGGATGCAGATTCCTGATTCAGTGCGGCCCGGAGAGAGTG 180
QY 1019 GTCACCGTCCGAGTCCCGACATCATGAGGAAGGATCATACCTTATTTTGGGAATTTGGCCACA 1078
DB 181 GTCAGTGTTCAGTATCCACCCATGAGAGAGGATCATATCTCTCTCGAATTTGCCACA 240
QY 1079 GACAGTTATGACATTTGGGTTTATTTTGAATGGACAGACTCTTCCAAATGCTGCT 1138
DB 241 GACAATTTATGACATTTGGGTTTATTTTGAATGGACAGACTCTTCCAAACACTGCT 300
QY 1139 CTCAGTGTGCATGTTCAGTGTCCAGTCCAGTGCACGAGGAGGAGGAGGAGGAGGAGGAGGAG 1198
DB 301 GTCAGCGTGCATGTTCAGTGTCCAGTGCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
```





1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
e mouse tissues.

## FEATURES

```

1. .493
location/qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7420465P14"
/clone_lib="RIKEN full-length enriched, in vitro
fertilized eggs"
/sex="female"
/tissue_type="in vitro fertilized eggs"
/dev_stage="egg"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATTCGACAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using therhalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTTCGAGTTAATTAATTCCTCCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
159 a 102 c 125 q
COUNT

```

Query Match	33.3%;	Score 485.6;	DB 10;	Length 493;
Best Local Similarity	99.2%;	Pred. No. 1.8e-104;		
Matches 488;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	709	ACAACAGGCAGCATTCACAGAAACAGCAAGAAGTAGTGATGCTGGGGCATCATTTGCCCTGC	768	
Db	1	ACAACAGCCAGCATTCACAGAAACAGCAAGAAGTAGTGATGCTGGGGCATCATTTGCCCTGC	60	
QY	769	ATCATCAAAAGGTGAACACAGCTGGAGCAAGTGTATACACTGTCAGTTAATGCACAGGCCAA	828	
Db	61	ATCATCAAAAGGTGAACACAGCTGGAGCAAGTGTATACACTGTCAGTTAATGCACAGGCCAA	120	
QY	829	AACCCACACTGAAATTCGCCAAAAGTCCCTTGAGCCACAGAAGCTGCACAGAAAGCCCTTGA	888	
Db	121	AACCCACACTTAAATTCGCAAAAAGTCCCTTGAGCCACAGAAGCTGCACAGAAAGCCCTTGA	180	
QY	889	AAATGACCACAAAAGACTCTCTCCAGTGATTTGCAGGCTCCCATCATGTGGACAAGACCACA	948	

Db	181	AAATGGACCAAAAAACATCTCTTCAGTGATTCAGCTCCATTCAGTGGACAGACCA	240
QY	949	AATCAAAAGACTTTAAAGAGAAAGATTTCGGCAGGATGCAGATTCTGTGATTACAGTACGTGC	1008
Db	241	AATCAAAAGACTTTAAAGAGAAATTCGGCAGGATGCAGATTCTGTGATTACAGTACGTGC	300
QY	1009	AGGAGAAGTCGTACCCGTCGAGTCCCGACTCATGAGGAAGGATCATACCTATTTTGGGA	1068
Db	301	AGGAGAAGTCGTACCCGTCGAGTCCCGACTCATGAGGAAGGATCATACCTATTTTGGGA	360
QY	1069	ATTGGCCACAGACAGTTATGACATTGGGTTTGGGTTTATTTCGAATGCAGACTCTCC	1128
Db	361	ATTGGCCACAGACAGTTATGACATTGGGTTTGGGTTTATTTCGAATGCAGACTCTCC	420
QY	1129	AAATGCTGCTGTCAGTGTGCATGTCAGTGCAGTGACGAGGAGGAGGAGGAAGA	1188
Db	421	AAATGCTGCTGTCAGTGTGCATGTCAGTGCAGTGACGAGGAGGAGGAGGAAGA	480
QY	1189	AAATGTCACCTTG 1200	
Db	481	AAATGTCACCTTG 492	
RESULT 8			
LOCUS	BI990454	600 bp mRNA linear	EST 20-DEC-2001
DEFINITION	4071-42 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,		
ACCESSION	BI990454	mRNA sequence.	
VERSION	BI990454.1	GI:17961466	
KEYWORDS	EST.		
SOURCE	house mouse		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 600)		
JOURNAL	White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.,		
MEDLINE	Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,		
COMMENT	Gene expression in the developing mouse retina by EST sequencing		
	and microarray analysis		
	Nucleic Acids Res. 29 (24), 4983-4993 (2001)		
	21671825		
	Contact: Klein WH		
	Department of Biochemistry and Molecular Biology		
	University of Texas M.D. Anderson Cancer Center		
	Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA		
	Tel: 713 792 3646		
	Fax: 713 790 0329.		
FEATURES	Location/Qualifiers		
SOURCE	1..600		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone_lib="Mouse E14.5 retina lambda ZAP II Library"		
	/tissue_type="neural retina"		
	/dev_stage="embryonic day 14.5 post-fertilization"		
	/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.		
	Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps		
	(Mammals); Cloning Technique: CUA Cloning (CloneAmp,		
	Life Technologies); Average insert size: 1.8 Kb;		
	Insertion site: TACGTCACTGAATCTGAGTG---. Other		
	information regarding entire library may be found at		
	http://pga.swmed.edu/data/libraries/microarray_cdna_library.htm."		
BASE COUNT	172 a 113 c 217 g 97 t	1 others	
ORIGIN			
Query Match	31.0%;	Score 452.4;	DB 13; Length 600;
Best Local Similarity	99.8%;	Pred. No. 1.3e-96;	
Matches 464; Conservative	0;	Mismatches 1;	Indels 1; Gaps 1;
QY	97	AGTAAAGATGGCAAGCCCTTCATCCACTTATGAGAAAACTGAAGTTCGTGGCACT	156



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999). Contact: Robert Strausberg, Ph.D. Email: c9apbs-r@mail.nih.gov Tissue Procurement: Louis Staudt, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9552 row: f column: 04 High quality sequence stop: 702.		REFERENCE AUTHORS TITLE JOURNAL COMMENT	
BASE COUNT 220 a 139 c 197 g 93 t			
ORIGIN			
Query Match 30.5%; Score 445.4; DB 12; Length 649; Best Local Similarity 92.5%; Pred. No. 6e-95; Matches 602; Conservative 0; Mismatches 31; Indels 18; Gaps 12;			
Qy	314 TGTGTCTCTCCTCGGATATGTTGGTCCACAGATAGAGGAAGGAAGAG 373		
Db	1 TGTGGTCTCTCTCGCATATG-TGCGTCCACAGATAGAGAAGGAAGAAG 59		
Qy	374 AAA-AGAGAAGCGGAGGAGCGAAGCGAGCTGAAGAGGAAGACGAGCGCT 432		
Db	60 AAAGAGAAGAGCGGAGGAGCGAAGCGAGCTGAAGAGGAAGACGAGCGGCT 119		
Qy	433 GCAAAGGAAGAGAGCGGAAGCG-ACAGGAGGAAGACCGGCTGAGACGGAGGAG 491		
Db	120 GCAAAGGAGNAGAGCGGAAGCGCAGAGGAGNAGACCGGCTGAGACGGAGGAG 179		
Qy	492 AAGAGGCGCGATAGAGGAAGAGAGGCTTCGGCTGGAACAGCAAAAGCAGAGATA 551		
Db	180 AAGAGGCGCGGATAGAGGAAGAGAGGCTTCGGCTGGAACAGCAAAAGCAGAGATA 239		
Qy	552 TGGCAGCTTTAACTCCAGA--CTGCCGTGCAATTCAC-AGTATGACGCCAGC-AG 607		
Db	240 TGGCAGCTTTAACTCCGAAAGCTGCCGTGCAATTCACGCGAGTATGACGCCAGCAAG 299		
Qy	608 TATCAGGGAACACGAACACAGC--AGATTCTCATTCCGCCAGCTGC-AGGAGCAG-CA 663		
Db	300 TATCAGGGAACACGAACACAGCAGGATTTCTATCCGCCAGCTGCGAGGAGCAGTCA 359		
Qy	664 CTATCAGCAGTATAAACACAGGCGAGCAAAACCCACCTGCAACAAC---CAGGCGAG 719		
Db	360 CTATCAGCAGTATATGCACGCGAGTTGATATCAAGTCCAGCTTGCACAAACAGCGCAG 419		
Qy	720 CATTACAGAACACAGAGAGTAGTAGTGCTGGGGCATCATTTGCCGTGCATCATCAAGG 779		
Db	420 CATTACAGAACACAGAGAGTAGTAGTGCTGGGGCATCATTTGCCGTGCATCATCAAGG 479		
Qy	780 TGAACACAGCTGGAGCAAGTGATACACTG- -CAGTTAATGGACAGGCCCAAAACCCACAC 837		
Db	480 TGAACACAGCTGGAGCAAGTGATACACTGTGCAGTGTAAATGGACAGGCCCAAAACCCACAC 539		
Qy	838 TGAATAATCCGAAAAAGTCTTGGCCAGAGAGCTGCAGAGAAGAGCCCTTGGAAAAATGGACC 897		
Db	540 TGAATAATCCGAAAAAGTCTTGGCCAGAGAGCTGCAGAGAAGAGCCCTTGGAAAAATGGACC 599		
Qy	898 AAAAGACTCTCTCCAGTGAATGTCAGCTCCATCCATGTCAGGAGACCCACA 948		
Db	600 AAAAGACTCTCTCCAGTGA-TGCAGCTCCATCCATGTCAGGAGACCCACA 649		
RESULT 11			
BF797451			
LOCUS 60257111F1 NIH_MGC_85 984 bp mRNA linear EST 12-JAN-2001			
DEFINITION mRNA sequence.			
ACCESSION BF797451			
VERSION BF797451.1 GI:12102505			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			

```

QY 1446 GTTTAAATTT 1454
Db 540 ATTAAATTT 548

RESULT 12
BE694778
LOCUS BE694778 519 bp mRNA linear EST 11-SEP-2000
DEFINITION PMI-BT0759-200700-006-c03 BT0759 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE694778
VERSION BE694778.1 GI:10081938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 519)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-PM1-BT0759-200700-006-c03&t3=2000-07-20&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 519.
Location/Qualifiers
1..519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0759"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 160 a 110 c 138 g 111 t
ORIGIN
Query Match 27.8%; Score 405; DB 12; Length 519;
Best Local Similarity 88.0%; Pred. No. 2.2e-85;
Matches 441; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 845 TCCGAAAAGTCCTTGAGCGCAGAGCTGCAGAAAGCCCTTGAAATCGACCAAAAGAC 904
Db 19 TCCACACAGAACTGGAACCGACAGCTGCAGAGAGCCCTGGAGATCGACCAAGAA 78
QY 905 TCTCTTCAGTGTGACGCTCCATCCATGTGGACAGACCAACAATCAAGACTTTAA 964
Db 79 TCTCTCCAGTAATAGCAGCTCCATCCATGTGGACAGACCTCGATCAAGACTTCAA 138
QY 965 GAGAAATTCGGCAGGATCGCATCTGTGATTACAGTACGTGAGGAGAACTGTCACC 1024
Db 139 GAGAGATTTCAGCAGGATCGAGATTCGCTGATTACAGTGGCGGAGAGAGTGGTCACT 198

```

```

QY 1025 GTCCGAGTCCCGACTCATGAGGAAGGATCATACCTATTATTGGGAATTTGCCACAGACAGT 1084
Db 199 GTTCGAGTACCACCCATGAAGAAGGATCATATCTTTTGGGAATTTGCCACAGACAAT 258
QY 1085 TATGACATGGGTTGGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTCAGT 1144
Db 259 TATGACATGGGTTGGGGTGTATTTTGAATGGACAGACTCTCCAAACACTGCTGTCAGC 318
QY 1145 GTGCATGTCACTGAGTCCAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1204
Db 319 GTGCATGTCACTGAGTCCAGCGATGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 378
QY 1205 GAAAAGCAAAAAGAACGCCCAACAGCCTCTGCTGGATGAGATTGTCTGTTGTACCGG 1264
Db 379 GAGAAAGCAAAAAGAAATGCCAACAAAGCCTTTCCTGGATGAGATGGTCTGTTGTACCGA 438
QY 1265 CGGGACTGTCAGGAGGAAGTATATGCAGCAGCCACCAGTATCCAGGGAGGGAGGAGTCTAT 1324
Db 439 CGGGACTGTCATGAGGAGGAGTGTATGCTGCAGCCATCAATATCCAGGGAGGAGGAGTCTAT 498
QY 1325 CTCCTCAAGTTTGTATAATTCC 1345
Db 499 CTCCTCAAGTTTGCACAACCTCC 519

RESULT 13
LOCUS BQ180694
DEFINITION UI-M-EX0-bxc-c-15-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
IMAGE:5706182 5', mRNA sequence.
ACCESSION BQ180694
VERSION BQ180694.1 GI:20356186
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 682)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, Univeristy of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 91-227, >GC-richFlow_complexity (matched complement)
548-664, >(GAA)n$Simple_repeat
Seq primer: pyx-5.
Location/Qualifiers
1..682
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5706182"
/clone_lib="NIH_BMAP_EX0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pyx-Asc; Site_1: EcoR I;
Site_2: Not I. The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded

```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 639)  
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

FEATURES source

```

location/Vadalliers
1. 639
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4700208"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site1:
SfiI (ggccgctcgcc); Site2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGCCGACATC-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

```

BASE COUNT	225 a	140 c	169 q	105 t
------------	-------	-------	-------	-------

## ORIGIN

Query Match 25.2%; Score 367.4; DB 12; Length 639;  
Best Local Similarity 84.1%; Pred. No. 1.8e-76;  
Matches 497; Conservative 0; Mismatches 86; Indels 8; Gaps 7;

QY 416 GAAGAACGAGCGGCTGCAAAAGGAAGAGAGAAGCGGAGAGGAGGAGACCGG 475

1b 3 GAAGAAAGAGAACGTCTGC AAAAGGAGGAAGAGAAACGTACGACAGACGAAGAGGAAAGG 62

476 CTGAGACGGGAGGGAAGAGAGGGCGGCGGATAGAGGAAGAGAGGCTTCGGCTGGAAACAG 535

Db 63 CTTGACGGGAGGAGGAAGGACCGGATAGAAGAAGAAAGGCTTCGTTGGAGCAG 122

Qy 536 CAAAAGCAGCAGATAATGGCAGCTTTAAACTGCAGACTGCCGTGCAATTCACGAGTAT 595

Db 123 CAAAAGCAGCAGATAATGGCAGCCTTTAAACTCCCAGACTGCCGTGCAGTTCACAGCAGTAT 182

QY 596 GCAGCCCGCAGTATCCAGGGAACTACGAACAACAGCAGATTCTCATCCGCCAGCTGCAG 655

Db 183 GCAGCCCAACAGTATCCA -GGAACACGACGAGCAAAATTCATCGCAGTTGCAG 241

QY 656 GAGCAGCACTATCAGCAGTATAAACACCCAGGCAG-AGCAAAACCCACCTGGCACAAACA 714

Db 242 GAGCAACACTATCAGCAGTACATGCACGAGTTGGTATCAAGTCCAGCTTGCACAGCAACA 301

Qy 715 GGCAGCATTACAGAAACAGCAAGAAGTAGTGATGGCTGGGGCATCATTCGCTGCATCATC 774

Db 302 GGCAGCATTACAGAAACACAGGAAGTAGTAGTGGGTCTTCCCTTGGCTACATCATC 361

775 AAAGGTGAACACACAGCTGGAGCAAGTGATACACTGTCAGTTAATGGACAGGCCAAAACCCA 834

DD AAAAGTGAATGCAACTGTACCAAGTAATATGAGTTCAGTTAATGGACAGGCCAAAACACA 421

[illegible]

004 THURGOOD PROCTOR UNIVERSITY COLLEGE OF BUSINESS AND ECONOMICS FEB 22

Qy 892 TGGACCAAAGACTCTCTCCAGTGATTGCAGCTCCAT - CCATGTGGACAGACCACA - A 949

Db 482 GGACCAACAAGATCTCTTCAGTAATAGCAGCTCCATCCATGTGGACACGACCTCAGA 541

Qy 950 ATCAAGACTT - TAAAGAGAAGATTCCGCGAGGATGCAGATTCCTGTGATTAC 999

Db 542 ATCAAGACTTCAAAGAGAAGATTTCAGCAGGATGCAGATTCGCGATTTC 592

Search completed: February 21, 2003, 23:27:09  
Job time : 2239 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 21:31:46 ; Search time 3906 Seconds  
(without alignments)  
10870.710 Million cell updates/sec

Title: US-09-762-594-2  
Perfect score: 1459  
Sequence: 1 gaattcgccgcgcgcac.....catgaagtttaatttccttt 1459

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1340.6	91.9	1543	10	AF022770	AF022770 Mus muscu
2	1079.6	74.0	2140	9	AK025520	AK025520 Homo sapi
3	1079.6	74.0	3598	9	AB043587	AB043587 Homo sapi
4	890.4	61.0	3049	9	BC034563	BC034563 Homo sapi
5	287.2	19.7	18140	10	AF501319	AF501319 Mus muscu
6	246.2	16.9	63971	2	AC121292	AC121292 Mus muscu
c 7	238	16.3	140409	2	AC044825	AC044825 Homo sapi
c 8	238	16.3	181719	2	AL592045	AL592045 Homo sapi
c 9	238	16.3	232180	2	AC021883	AC021883 Homo sapi
c 10	228.4	15.7	174143	2	AC126290	AC126290 Rattus no
c 11	211.4	14.5	2235	9	AK057118	AK057118 Homo sapi
c 12	204.4	14.0	934	6	AX339070	AX339070 Sequence
c 13	158	10.8	174143	2	AC126290	AC126290 Rattus no
c 14	150.8	10.3	3547	9	AK095650	AK095650 Homo sapi
c 15	150	10.3	882	6	AX400067	AX400067 Sequence
c 16	113.4	7.8	527	6	AX339069	AX339069 Sequence
c 17	110.6	7.6	210204	9	AC007954	AC007954 Homo sapi
c 18	102.2	7.0	227588	2	AC094571	AC094571 Rattus no
c 19	100.8	6.9	1779	3	AY051848	AY051848 Drosophil
c 20	99.8	6.8	150641	3	CEY41E3	295559 Caenorhabdi
c 21	99.6	6.8	7218	6	I66494	I66494 Sequence 14
c 22	95.6	6.6	155662	2	AC112388	AC112388 Rattus no
c 23	95.6	6.6	189461	2	AC119627	AC119627 Rattus no
c 24	95.6	6.6	191841	2	AC121374	AC121374 Rattus no
c 25	95.2	6.5	84472	2	AC096684	AC096684 Takifu
c 26	95	6.5	264522	2	AC090437	AC090437 Mus muscu
c 27	94	6.4	204259	2	AC110817	AC110817 Mus muscu
c 28	93.8	6.4	184865	2	AL807824	AL807824 Mus muscu
c 29	93.6	6.4	177035	2	AC099443	AC099443 Rattus no
c 30	93.4	6.4	190594	10	AL606965	AL606965 Mouse DNA
c 31	93.2	6.4	170985	2	AC096032	AC096032 Rattus no
c 32	92.6	6.3	225045	2	AL833803	AL833803 Mus muscu
c 33	92.4	6.3	176030	2	AC129650	AC129650 Rattus no
c 34	92.4	6.3	287058	2	AC095595	AC095595 Rattus no
c 35	92.2	6.3	195673	2	AL807807	AL807807 Mus muscu
c 36	91.8	6.3	180418	2	AC106674	AC106674 Rattus no
c 37	91.8	6.3	272545	2	AC090533	AC090533 Mus muscu
c 38	91.4	6.3	131346	2	AC119558	AC119558 Rattus no
c 39	91.4	6.3	197247	2	AC125713	AC125713 Rattus no
c 40	91	6.2	261604	2	AC119819	AC119819 Mus muscu
c 41	90.2	6.2	168210	2	AC116700	AC116700 Mus muscu
c 42	90	6.2	82400	10	AC090495	AC090495 Genomic s
c 43	89.8	6.2	186559	10	AL606742	AL606742 Mouse DNA
c 44	89.6	6.1	107637	2	AC094679	AC094679 Rattus no
c 45	89.2	6.1	130580	2	AC119765	AC119765 Rattus no

ALIGNMENTS

RESULT 1  
AF022770  
LOCUS  
DEFINITION Mus musculus peripheral benzodiazepine receptor associated protein  
1543 bp mRNA linear ROD 11-APR-2002  
(Pap7) mRNA, complete cds.  
ACCESSION AF022770  
VERSION AF022770.2 GI:7545290  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1543)  
AUTHORS Li,H., Degenhardt,B., Tobin,D., Yao,Z.X., Tasken,K. and Papadopoulos,V.

**TITLE** Identification, localization, and function in steroidogenesis of PAP7: a peripheral-type benzodiazepine receptor- and PKA (R1alpha)-associated protein

**JOURNAL** Mol. Endocrinol. 15 (12), 2211-2228 (2001)

**MEDLINE** 21588728

**PUBMED** 11731621

**REFERENCE** 2 (bases 736 to 1290)

**AUTHORS** Li.H. and Papadopoulos.V.

**TITLE** Direct Submission

**JOURNAL** Submitted (03-SEP-1997) Department of Cell Biology, Georgetown University Medical Center, 3900 Reservoir Rd. NW, Washington, DC 20007, USA

**REFERENCE** 3 (bases 1 to 1543)

**AUTHORS** Li.H. and Papadopoulos.V.

**TITLE** Direct Submission

**JOURNAL** Submitted (12-APR-2000) Department of Cell Biology, Georgetown University Medical Center, 3900 Reservoir Rd. NW, Washington, DC 20007, USA

**REMARK** Sequence update by submitter

**COMMENT** On Apr 12, 2000 this sequence version replaced gi:2432020.

**FEATURES** Location/Qualifiers

1..1543

/organism="Mus musculus"

/strain="BALB/c"

/db\_xref="taxon:10090"

1..1543

/gene="pap7"

/gene="pap7"

101..1438

/note="PBR associated protein; PAP7"

/codon\_start=1

/product="peripheral benzodiazepine receptor associated protein"

/protein\_id="BAB71197.2"

/db\_xref="GI:7545291"

/translation="MRFRRAAEVRVKDKAFHPTVEELKLFVALHKQVLGYPND TSPVGFVLGNDRRFEALGNMSDEAMVEFKLLNKCPLLSAYVASHRIKEE EKKRKAERREERERLQKEEKRRREEDRLRREERERRRTEERLRLEQOK QOIMALNSTAVFOQYAAQYQYGNVEQOQILRLQEOHYQYMOQLXVOLAQOQ AALQKQEVYMAASLSPASSKVNTAGSDRLSVNGQAKTHENSEKYLEPEAREALE NGPKDSPIVIAAPSMWRPQIKDFKIRQDADSVITVRKGEVVTVRPHEEGSYLF WEFATDSDYIGFYFFWTDSPNAVSVHVSSEDEEEENVTCEKAKKNANKPL LDEIVPYRDCHEVYAGSHQYPGRGVILLKFDNSYSLWRSKSVYVYVYR"

BASE COUNT 484 a 310 c 437 g 312 t

ORIGIN

Query Match 91.98; Score 1340.6; DB 10; Length 1543;

Best Local Similarity 99.04; Pred. No. 0;

Matches 1349; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 97 AGTAAAGATGCAAGCCTTTTCATCCAACTTATGAAGAAAGAACTGAAGTTCGTGGCACT 156

DB 136 AGTAAAGATGCAAGCCTTTTCATCCAACTTATGAAGAAAGAACTGAAGTTCGTGGCACT 195

QY 157 GCATAAGCAAGTCTTTTGGGCCATATAACCCACACACGTCCTCGAGTTGGATTCTT 216

DB 196 GCATAAGCAAGTCTTTTGGGCCATATAACCCACACACGTCCTCGAGTTGGATTCTT 255

QY 217 TGATGTGTGGGAATGATAGAGAGAGATGGGCACCTCTGGGAACATGTCCAAGGA 276

DB 256 TGATGTGTGGGAATGATAGAGAGAGATGGGCACCTCTGGGAACATGTCCAAGGA 315

QY 277 GGATCCCATGTTAGAGTTGTGAAGCTTCTAAATGAAGTGTGCTCTCTCCCTCGGCATA 336

DB 316 GGATCCCATGTTAGAGTTGTGAAGCTTCTAAATGAAGTGTGCTCTCTCTCTCGGCATA 375

QY 337 TGTTCGCTCCACAGATAG 396

DB 376 TGTTCGCTCCACAGATAG 435

QY 397 GCGAAGGAGCGTGAAG 456

DB 436 GCGAAGGAGCGTGAAG 495

QY 457 GCGAGAGAGGAAGACCGGCTGAGACGGGAGGAGGAAGAGAGCGCGGATAGAGAAGA 516

DB 496 GCGAGAGAGGAAGACCGGCTGAGACGGGAGGAGGAAGAGAGCGCGGATAGAGAAGA 555

QY 517 GAGGCTTCGGCTGGAACAGCAAGCAAGCAGCAGATATATGCAGCTTTAAACTCGCAGCTGC 576

DB 556 GAGGCTTCGGCTGGAACAGCAAGCAAGCAGCAGATATATGCAGCTTTAAACTCGCAGCTGC 615

QY 577 CGTCAATTCAGCAGTATGAGCCCGCAGCAGTATCCAGGGAACTACGAAACACACAGAT 636

DB 616 CGTCAATTCAGCAGTATGAGCCCGCAGCAGTATCCAGGGAACTACGAAACACACAGAT 675

QY 637 TCTCATCCCGCAGCTGCGAGGAGCAGCAGTATCAGCAGTATATAACACACGAGGAGCAAG 696

DB 676 TCTCATCCCGCAGCTGCGAGGAGCAGCAGTATCAGCAGTATATGCGAGCAGTATATCAAGT 735

QY 697 TCAACCTCCACAACAACAGCAGCAGCAGTATCAGAAAGTTCCTTGAGCCAGAGCTGCAGA 756

DB 736 CCAGCTTGCACAACAACAGCAGCAGCAGTATCAGAAAGTTCCTTGAGCCAGAGCTGCAGA 795

QY 757 ATCATTTGCTGCATCATCAAGGTTGAACACAGCTGGAGCAAGTGTATACACTGTCTAGTTAA 816

DB 796 ATCATTTGCTGCATCATCAAGGTTGAACACAGCTGGAGCAAGTGTATACACTGTCTAGTTAA 855

QY 817 TGGACAGGCCCAAAACCCACACTGAAATTTCCGAAAAAGTCTCTTGAGCCAGAGCTGCAGA 876

DB 856 TGGACAGGCCCAAAACCCACACTGAAATTTCCGAAAAAGTCTCTTGAGCCAGAGCTGCAGA 915

QY 877 AGAAGCCTTGGAAATGAGCAAAAGACTCTCTTCCAGTGTATTCGAGCTCCATCCATGTG 936

DB 916 AGAAGCCTTGGAAATGAGCAAAAGACTCTCTTCCAGTGTATTCGAGCTCCATCCATGTG 975

QY 937 GACAGACCAAAATCAAGACTTTAAAGAGAGATTCGGCAGGATGCAGATCTCTGTGAT 996

DB 976 GACAGACCAAAATCAAGACTTTAAAGAGAGATTCGGCAGGATGCAGATCTCTGTGAT 1035

QY 997 TACAGTACCTCGAGGAGAGTCTGTCACCGTCCGAGTCCCGACTCATGAGGAAGATCATA 1056

DB 1036 TACAGTACCTCGAGGAGAGTCTGTCACCGTCCGAGTCCCGACTCATGAGGAAGATCATA 1095

QY 1057 CCTATTTTGGAAATGCGCAGACAGATTTACATTTGGGTTTGGGTTTATTTTGAATG 1116

DB 1096 CCTATTTTGGAAATGCGCAGACAGTATACATTTGGGTTTGGGTTTATTTTGAATG 1155

QY 1117 GACAGACTCTCAAAATGCTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTG 1176

DB 1156 GACAGACTCTCAAAATGCTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTG 1215

QY 1177 GGAGGAGGAAGAAATGTCACCTTGTGAAGAAAGCAAAAGCAAGCAAGCAAGCAAGCTCT 1236

DB 1216 GGAGGAGGAAGAAATGTCACCTTGTGAAGAAAGCAAAAGCAAGCAAGCAAGCAAGCTCT 1275

QY 1237 GCTGGATGAGATTTGTCACCTGTGTACCGCGGAGCTGTCCAGGAGAGATATATGAGGAG 1296

DB 1276 GCTGGATGAGATTTGTCACCTGTGTACCGCGGAGCTGTCCAGGAGAGATATATGAGGAG 1335

QY 1297 CCACAGTATCCAGGAGGAGGAGTCTATCTCTCAAGTTTGTATATTCCTACTCTCTGTG 1356

DB 1336 CCACAGTATCCAGGAGGAGGAGTCTATCTCTCAAGTTTGTATATTCCTACTCTCTGTG 1395

QY 1357 GAGGTCCCAAGTCCGCTCTACTACAGAGTCTATATATAGATAGAGCTGCTGTCTTCCAAGT 1416

DB 1396 GAGGTCCCAAGTCCGCTCTACTACAGAGTCTATATATAGATAGAGCTGCTGTCTTCCAAGT 1455

QY 1417 CCGGAGTCCAGGTTGAGCAACAACATGAGTTTAAATTCCTTT 1459

DB 1456 CCGGAGTCTAGGTTGAGCAACAACATGAGTTTAAATTCCTTT 1498

RESULT 2

AK025520

LOCUS

AK025520 2140 bp mRNA linear PRI 29-SEP-2000





```
QY 1237 GCTGGATGAGATTGACCTGTGTACCGGGGAGCTGTACGAGGAGAGTATATGTCAGGCAG 1296
Db 1480 GCTGGATGAGATTGCTGTGTACCGAGCGGACTGTATGAGGAGGTGTATCTCTGGCAG 1539
QY 1297 CCACGAGTATCAGGAGGAGGAGTCTATCTCTCAAGTTTGATTAATTCCTACTCTCTGTG 1356
Db 1540 CCATCAATATCAGGAGGAGGAGTCTATCTCTCAAGTTTGATTAATTCCTACTCTCTGTG 1599
QY 1357 GAGGTCCAAGTCCGCTACTACAGAGTCTATTATATCTAGATAGAGTGTCTTCCAAAGT 1416
Db 1600 GGGTCAAAATCAGTCTACTACAGAGTCTATTATCTAGATAAAATGTTGTTACAAGT 1659
QY 1417 CCGGAGTCAGGGTTCAGCACACATGAGCTTTAATTT 1454
Db 1660 CTGGAGTCTAGGTTGGCAGAGATGACATTTAATTT 1697

RESULT 4
BC034563 3049 bp mRNA linear PRI 26-JUL-2002
DEFINITION Homo sapiens, similar to golgi complex associated protein 1, 60
kDa, clone IMAGE:3858463, mRNA, partial cds.
ACCESSION BC034563
VERSION BC034563.1 GI:21961496
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 3049)
Strausberg,R.
Direct Submission
Submitted (24-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabps@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 21 Row: c Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
1. 3049
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3858463"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_66"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
<1..1095
/codon_start=1
/product="Similar to golgi complex associated protein 1,
60 kDa"
/protein_id="AAH34563.1"
/db_xref="GI:21961497"
/translation="RTRGCHLFSTYVASHKTEKEFEQKKRKEEERREERLERLQ
```

```
KEEKRRREERLRREERREERLRLEEQKQKQIMALNSQTAVQFOQYAAQY
PGNYEQIILIRLOEQHYQYMOQLYQVLAQQAALQKQEPYVAVAGSLPTSSKYN
ATVPSNMNSVNGQAKTHTDSSEKLEPEAEALENGPKESLPYAAPSMWTPOIKD
FKEXIQDADSIVTVGRGEVVTVRVPTHEGSLYLFWEFATDNYDIGFVFEWTDSPN
TAVSVHYSESDDDEBEENIGCEKAKKNANKPLLEIYVVPYRRDCHEEYVAGSHQY
PGRVYLLKFDNSLSLWRSKSVYIRVYITR*
BASE COUNT 963 a 522 c 633 g 931 t
ORIGIN
Query Match 61.0%; Score 890.4; DB 9; Length 3049;
Best Local Similarity 86.3%; Pred. No. 8.1e-216;
Matches 984; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
QY 315 GTTGTCTCTCTCTCGGCATATGTTGGTCCACAGATAGAGGAAGAGAGAGA 374
Db 11 GGTGCCATCTCTTCAACATATGTTGGTCCCAAAATAGAGAAGAGCAAGAA 70
QY 375 AAAGAAGAAAGCGGAGGAGGAGCGGAGCGGTGGAAGAGGAAGAACGAGAGCGGTGC 434
Db 71 AAAAAGGAAGGAGGAGGAGGAGCGGAGCGGTGGAAGAGGAAGAGAACGTCGTGC 130
QY 435 AAAAGGAAGAGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 494
Db 131 AAAAGGAGGAGAGAAACGTAGGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 190
QY 495 AGAGCGCGCGGATAGAGGAAGAGAGGCTTGGCTGGAACAGCAAAAGCAGCAGATAATGG 554
Db 191 AAAGCAGCGGATACAGAAAGAAAGCGTTCGGTGGAGCAGCAAAAGCAGCAGATAATGG 250
QY 555 CAGCTTTAACTCGCAGACTGCCGTGCAATTCACGAGATATCCAGCCAGCAGCAGTATCCAG 614
Db 251 CAGCTTTAACTCCAGACTGCCGTGCAATTCACGAGATATCCAGCCAGCAGCAGTATCCAG 310
QY 615 CGAATACCAACAGCAGAGATCTCATCCGCCAGCTGCAGGAGCAGCAGCTATCAGCAGT 674
Db 311 GGAATACCAACAGCAGCAAAATTCATCCGCCAGTTCGAGGAGCAACACTATCAGCAGT 370
QY 675 ATAAACACGAGCAGCAAGCAAAACCCCACTGCACACACAGCAGCAGCTATACAGAAACAGC 734
Db 371 ACATCAGCAGTGTATCAAGTCCAGCTTGCACACACAGCAGCAGCTATACAGAAACAC 430
QY 735 AAGAAGTAGTGTGCTGGGCGATCATGTCCTGTCATCATCAAGGTGAACACAGCTGGAG 794
Db 431 AGGAAGTAGTGTGCTGGGCTCTCTCTTGCCTACATCATCAAAAGTGAATGCAACTGATC 490
QY 795 CAAGTGATACACTGTCAGTTAATGACAGCGCCAAACCCACACTGAAATTCGGAAGAG 854
Db 491 CAAGTAATATGATGTCAGTTAATGACAGCGCCAAACACACTGACAGCTCCGAAAGAG 550
QY 855 TCCTTGAGCCAGAAGCTGCAGAAAGAGCGCTTGGAAATGGACCAAAAGACTCTCTTCCAG 914
Db 551 AACTGGAACCAAGCTGCAGAAAGAGCGCTTGGAAATGGACCAAAAGACTCTCTTCCAG 610
QY 915 TGATTCAGCTCCATCCATGTGGACAAGACACAAATCAAGACTTTAAAGAGAGATTC 974
Db 611 TAATAGCAGCTCCATCCATGTGGACAGCAGCTCAGATCAAAAGACTTCAAAAGAGAGATTC 670
QY 975 GGCAGGATCAGATTCGTGATTACAGTACGTCGAGGAGAGTCTGTCACCGTCCGAGTCC 1034
Db 671 AGCAGGATCAGATTCGCGTGAATACAGTGGCGGAGGAGAGTGGTCACTGTTCAGATAC 730
QY 1035 CGACTCATGAGGAAGATCATATCTATTTTGGGAATTTGCCACAGACAGATTTATGACATTC 1094
Db 731 CCACCCATGAAGAAGATCATATCTCTTTGGGAATTTGCCACAGACAAATTTATGACATTC 790
QY 1095 GCTTTGGGTTTATTTTGAATGGACAGACTCTCCAAATGCTGCTGTGATGTCATGTCA 1154
Db 791 GCTTTGGGTTGATTTTGAATGGACAGACTCTCCAAACACTGCTGTGTCGCTGATGTCA 850
QY 1155 GTGAGTCCAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1214
Db 851 GTGAGTCCAGCATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 910
```

QY	1215	AAAAGACGCCACAAGCCTCTGCTGGATGAGATTGTACCTGTACCGGGGACTGTC	1274
Db	911	AAAAGAATGCCACAAGCCTTTGCTGGATGAGATTGTGCTGTACCGACGGACTGTC	970
QY	1275	ACGAGGAGATATATGCAGGACGCCACCAAGTATCAGGAGGGGAGCTATCTCTCAAGT	1334
Db	971	ATGAGGAGGTGTATGCTGGCAGCCATCAATATCAGGAGGAGGAGTCTATCTCTCAAGT	1030
QY	1335	TTGATAATTCTACTCTCTGTGGAGGTCCTCAAGTCCGCTCTACTACAGAGTCTATTATACTA	1394
Db	1031	TTGCACACTCTACTCTTTGTGGCGGTCAAAATCAGTCTACTACAGAGTCTATTATACTA	1090
QY	1395	GATAGAGCTGCTGTTCCAAAGTCCGGAGTCCAGGGTTGAGCACAACTGACGTTTAATTT	1454
Db	1091	GATAAAATGTTGTTACAAAGTCTGGAGTCTAGGGTTGGCAGAAGATGACATTTAATTT	1150
RESULT 5			
AF501319			
LOCUS	AF501319	18140 bp DNA linear	ROD 06-MAY-2002
DEFINITION	Mus musculus peripheral benzodiazepine receptor associated protein		
ACCESSION	PAP7 (PAP7) gene, complete cds.		
VERSION	AF501319		
KEYWORDS	AF501319.1 GI:20453988		
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 18140)		
TITLE	Liu, J and Papadopoulos, V.		
JOURNAL	Direct Submission		
	Submitted (11-APR-2002); Department of Cell Biology, Georgetown University Medical Center, 3900 Reservoir Rd. NW, Washington, DC 20007. USA		

[illegible]

D	b	15039	CCTTTAGACTCTCTTCACGTGATTTGCAGTCCATCCATCGCATGGACAAGACCACAATCAAA	15098
Q	y	956	GACTTTTAAAGAAGAATTGGCAGGATGCAGATTCTCTGATTACAGTAGCTCGAGGAGAAA	1015
D	b	15099	GACTTTTAAAGAAGAATTGGCAGGATGCAGATTCTCTGATTACAGTAGCTCGAGGAGAAA	15158
Q	y	1016	GTGCTCACCGTCCGAGTCCCGACTCATGAGGAAGGATCATACCTATTTTGGGAATTTGCC	1075
D	b	15159	GTGCTCACCGTCCGAGTCCCGACTCATGAGGAAGGATCATACCTATTTTGGGAATTTGCC	15218
Q	y	1076	ACAGACAGTATGACATTGGCTTTGGGGTTTATTTTGAATGGACAGACTCTCCAATATGCT	1135
D	b	15219	ACAGACAGTATGACATTGGCTTTGGGGTTTATTTTGAATGGACAGACTCTCCAATATGCT	15278
Q	y	1136	GCTCTCAGTGTCATGTCACTGACGTCCAGTGCACGAGGAGGAGGAGGAAG	1187
D	b	15279	GCTCTCAGTGTCATGTCACTGACGTGCACGAGGAGGAGGAGGAGGAAG	15330

RESULT 6  
AC121292                  63971 bp      DNA      linear      HTG 16-MAY-2002

LOCUS Mus musculus clone RP23-172B15, LOW-PASS SEQUENCE SAMPLING.  
AC121292

DEFINITION AC121292.1 GI:20806432

ACCESSION AC121292.1

VERSION HTG; HTGS\_PHASE0.

KEYWORDS Mus musculus.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 63971)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

AUTHORS Unpublished

TITLE 2 (bases 1 to 63971)

JOURNAL Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckhaider,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazarro,B., Choepel,A., Colangelo,M., Collins,P., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Gord,S., Goette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneses,L., Mihova,T., Miengua,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemlek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE Submitted (16-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL All repeats were identified using RepeatMasker:

COMMENT Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L24476  
Center clone name: 172\_B\_15

\* NOTE: This record contains 79 individual

\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 717: contig of 717 bp in length  
\* 718 817: gap of 100 bp  
\* 818 1503: contig of 686 bp in length  
\* 1504 1603: gap of 100 bp  
\* 1604 2317: contig of 714 bp in length  
\* 2318 2417: gap of 100 bp  
\* 2418 3133: contig of 716 bp in length  
\* 3134 3233: gap of 100 bp  
\* 3234 3959: contig of 726 bp in length  
\* 3960 4059: gap of 100 bp  
\* 4060 4780: contig of 721 bp in length  
\* 4781 4880: gap of 100 bp  
\* 4881 5585: contig of 705 bp in length  
\* 5586 5685: gap of 100 bp  
\* 5686 6398: contig of 713 bp in length  
\* 6399 6498: gap of 100 bp  
\* 6499 7202: contig of 704 bp in length  
\* 7203 7302: gap of 100 bp  
\* 7303 8010: contig of 708 bp in length  
\* 8011 8110: gap of 100 bp  
\* 8111 8835: contig of 725 bp in length  
\* 8836 8935: gap of 100 bp  
\* 8936 9652: contig of 717 bp in length  
\* 9653 9752: gap of 100 bp  
\* 9753 10487: contig of 735 bp in length  
\* 10488 10587: gap of 100 bp  
\* 10588 11289: contig of 702 bp in length  
\* 11290 11389: gap of 100 bp  
\* 11390 12087: contig of 698 bp in length  
\* 12088 12187: gap of 100 bp  
\* 12188 12887: contig of 700 bp in length  
\* 12888 12987: gap of 100 bp  
\* 12988 13698: contig of 711 bp in length  
\* 13699 13798: gap of 100 bp  
\* 13799 14510: contig of 712 bp in length  
\* 14511 14610: gap of 100 bp  
\* 14611 15322: contig of 712 bp in length  
\* 15323 15422: gap of 100 bp  
\* 15423 16103: contig of 681 bp in length  
\* 16104 16203: gap of 100 bp  
\* 16204 16929: contig of 726 bp in length  
\* 16930 17029: gap of 100 bp  
\* 17030 17755: contig of 726 bp in length  
\* 17756 17855: gap of 100 bp  
\* 17856 18577: contig of 722 bp in length  
\* 18578 18677: gap of 100 bp  
\* 18678 19391: contig of 714 bp in length  
\* 19392 19491: gap of 100 bp  
\* 19492 20201: contig of 710 bp in length  
\* 20202 20301: gap of 100 bp  
\* 20302 20993: contig of 692 bp in length  
\* 20994 21093: gap of 100 bp  
\* 21094 21812: contig of 719 bp in length  
\* 21813 21912: gap of 100 bp  
\* 21913 22605: contig of 693 bp in length  
\* 22606 22705: gap of 100 bp  
\* 22706 23426: contig of 721 bp in length  
\* 23427 23526: gap of 100 bp  
\* 23527 24238: contig of 712 bp in length  
\* 24239 24338: gap of 100 bp  
\* 24339 25037: contig of 699 bp in length  
\* 25038 25137: gap of 100 bp  
\* 25138 25858: contig of 721 bp in length

\* 25859 25958: gap of 100 bp  
\* 25959 26679: contig of 721 bp in length  
\* 26680 26779: gap of 100 bp  
\* 26780 27498: contig of 719 bp in length  
\* 27499 27598: gap of 100 bp  
\* 27599 28291: contig of 693 bp in length  
\* 28292 28391: gap of 100 bp  
\* 28392 29094: contig of 703 bp in length  
\* 29095 29194: gap of 100 bp  
\* 29195 29887: contig of 693 bp in length  
\* 29888 29987: gap of 100 bp  
\* 29988 30707: contig of 720 bp in length  
\* 30708 30807: gap of 100 bp  
\* 30808 31528: contig of 721 bp in length  
\* 31529 31628: gap of 100 bp  
\* 31629 32330: contig of 702 bp in length  
\* 32331 32430: gap of 100 bp  
\* 32431 33132: contig of 702 bp in length  
\* 33133 33232: gap of 100 bp  
\* 33233 33924: contig of 692 bp in length  
\* 33925 34024: gap of 100 bp  
\* 34025 34741: contig of 717 bp in length  
\* 34742 34841: gap of 100 bp  
\* 34842 35563: contig of 722 bp in length  
\* 35564 35663: gap of 100 bp  
\* 35664 36385: contig of 722 bp in length  
\* 36386 36485: gap of 100 bp  
\* 36486 37216: contig of 731 bp in length  
\* 37217 37316: gap of 100 bp  
\* 37317 38018: contig of 702 bp in length  
\* 38019 38118: gap of 100 bp  
\* 38119 38830: contig of 712 bp in length  
\* 38831 38930: gap of 100 bp  
\* 38931 39616: contig of 686 bp in length  
\* 39617 39716: gap of 100 bp  
\* 39717 40433: contig of 717 bp in length  
\* 40434 40533: gap of 100 bp  
\* 40534 41241: contig of 708 bp in length  
\* 41242 41341: gap of 100 bp  
\* 41342 42051: contig of 710 bp in length  
\* 42052 42151: gap of 100 bp  
\* 42152 42847: contig of 696 bp in length  
\* 42848 42947: gap of 100 bp  
\* 42948 43664: contig of 717 bp in length  
\* 43665 43764: gap of 100 bp  
\* 43765 44487: contig of 723 bp in length  
\* 44488 44587: gap of 100 bp  
\* 44588 45305: contig of 718 bp in length  
\* 45306 45405: gap of 100 bp  
\* 45406 46117: contig of 712 bp in length  
\* 46118 46217: gap of 100 bp  
\* 46218 46921: contig of 704 bp in length  
\* 46922 47021: gap of 100 bp  
\* 47022 47731: contig of 710 bp in length  
\* 47732 47831: gap of 100 bp  
\* 47832 48548: contig of 717 bp in length  
\* 48549 48648: gap of 100 bp  
\* 48649 49364: contig of 716 bp in length  
\* 49365 49464: gap of 100 bp  
\* 49465 50171: contig of 707 bp in length  
\* 50172 50271: gap of 100 bp  
\* 50272 50976: contig of 705 bp in length  
\* 50977 51076: gap of 100 bp  
\* 51077 51760: contig of 684 bp in length  
\* 51761 51860: gap of 100 bp  
\* 51861 52585: contig of 725 bp in length  
\* 52586 52685: gap of 100 bp  
\* 52686 53414: contig of 729 bp in length  
\* 53415 53514: gap of 100 bp  
\* 53515 54235: contig of 721 bp in length  
\* 54236 54335: gap of 100 bp  
\* 54336 55059: contig of 724 bp in length  
\* 55060 55159: gap of 100 bp

```
* 55160 55872: contig of 713 bp in length
* 55873 55972: gap of 100 bp

Query Match
Best Local Similarity 16.9%; Score 246.2; DB 2; Length 63971;
Matches 259; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1186 AGAAATGTCACCTGTGAAGAAAAGCAAAAGCAACGCGCCCTCTCTCGATGA 1245
Db 33662 AGAAATGTCACCTGTGAAGAAAAGCAAAAGCAACGCGCCCTCTCTCGATGA 33721

QY 1246 GATTGTACCTGTGTACCGGC-GGGACTGTACGAGGAAGTATATGAGCAGCCACCCAGT 1304
Db 33722 GATTGTACCTGTGTACCGCGGGGACTGTACGAGGAAGTATATGAGCAGCCACCCAGT 33781

QY 1305 ATCCAGGAGGAGGAGTCTATCTCCTCAAGTTTGAATTCCTACTCTCTGTGGAGTCCA 1364
Db 33782 ATCCAGGAGGAGGAGTCTATCTCCTCAAGTTTGAATTCCTACTCTCTGTGGAGTCCA 33841

QY 1365 AGTCCGCTACTACAGAGTCTATTATCTACTAGATAGAGTCTGTCCAGGTCGCGGAGTC 1424
Db 33842 AGTCCGCTACTACAGAGTCTATTATCTACTAGATAGAGTCTGTCCAGGTCGCGGAGTC 33901

QY 1425 CAGGGTTGAGCACACATGACGT 1447
Db 33902 TAGGGTTGAGCACACATGACGT 33924
```

```
RESULT 7
AC044825/c
LOCUS
DEFINITION
Homo sapiens chromosome 1 clone RP11-275114 map 1, WORKING DRAFT
SEQUENCE, 39 unordered pieces.
ACCESSION
AC044825.2 GI:9502452
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 140409)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 1, clone RP11-275114
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 140409)
AUTHORS
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McCurk,A., McKernan,K., McPheeters,R.,
McDrith,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Jul 27, 2000 this sequence version replaced gi:7543788.
All repeats were identified using RepeatMasker:
Smt,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 275_I_14
Center clone name: 18011
----- Summary Statistics
Sequencing vector: M13: W7815; 99% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
1.00182149362477Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 120391 bases at least Q40
Consensus quality: 129577 bases at least Q30
Consensus quality: 133768 bases at least Q20
Insert size: 160000; agarose-fp
Insert size: 136609; sum-of-contigs
Quality coverage: 3.1 in Q20 bases; agarose-fp
Quality co.
NOTE: This is a 'working draft' sequence. It currently
consists of 39 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1156: contig of 1156 bp in length
1157 1256: gap of 100 bp
1257 2265: contig of 1009 bp in length
2266 2365: gap of 100 bp
2366 3576: contig of 1211 bp in length
3577 3676: gap of 100 bp
3677 4845: contig of 1169 bp in length
4846 4945: gap of 100 bp
4946 7367: contig of 2422 bp in length
7368 7467: gap of 100 bp
7468 8595: contig of 1128 bp in length
8596 8695: gap of 100 bp
8696 10409: contig of 1714 bp in length
10410 10509: gap of 100 bp
10510 11838: contig of 1329 bp in length
11839 11938: gap of 100 bp
11939 13792: contig of 1854 bp in length
13793 13892: gap of 100 bp
13893 15478: contig of 1586 bp in length
15479 15578: gap of 100 bp
15579 17753: contig of 2175 bp in length
17754 17853: gap of 100 bp
17854 19331: contig of 1478 bp in length
19332 19431: gap of 100 bp
19432 21597: contig of 2166 bp in length
21598 21697: gap of 100 bp
21698 23136: contig of 1439 bp in length
23137 23236: gap of 100 bp
23237 25903: contig of 2667 bp in length
25904 26003: gap of 100 bp
26004 29651: contig of 3648 bp in length
29652 29751: gap of 100 bp
29752 33001: contig of 3250 bp in length
33002 33101: gap of 100 bp
33102 35597: contig of 2496 bp in length
35598 35697: gap of 100 bp
35698 37964: contig of 2267 bp in length
37965 38064: gap of 100 bp
38065 41070: contig of 3006 bp in length
41071 41170: gap of 100 bp
41171 43345: contig of 2175 bp in length
43346 43445: gap of 100 bp
43446 46364: contig of 2919 bp in length
46365 46464: gap of 100 bp
```







*	19474	20668:	contig	of 1195	bp in length
*	20669		gap	of unknown	length
*	20769	22399:	contig	of 1631	bp in length
*	22400		gap	of unknown	length
*	22500	24200:	contig	of 1701	bp in length
*	24201		gap	of unknown	length
*	24301	25667:	contig	of 1367	bp in length
*	25668		gap	of unknown	length
*	25768	27705:	contig	of 1938	bp in length
*	27706		gap	of unknown	length
*	27806	29643:	contig	of 1838	bp in length
*	29644		gap	of unknown	length
*	29744	31718:	contig	of 1975	bp in length
*	31719		gap	of unknown	length
*	31819	33299:	contig	of 1481	bp in length
*	33300		gap	of unknown	length
*	33400	35588:	contig	of 2189	bp in length
*	35589		gap	of unknown	length
*	35689	37968:	contig	of 2280	bp in length
*	37969		gap	of unknown	length
*	38069	40341:	contig	of 2273	bp in length
*	40342		gap	of unknown	length
*	40442	42574:	contig	of 2133	bp in length
*	42575		gap	of unknown	length
*	42675	44938:	contig	of 2264	bp in length
*	44939		gap	of unknown	length
*	45039	50126:	contig	of 5088	bp in length
*	50127		gap	of unknown	length
*	50227	52594:	contig	of 2368	bp in length
*	52595		gap	of unknown	length
*	52695	57156:	contig	of 4462	bp in length
*	57157		gap	of unknown	length
*	57257	61310:	contig	of 4054	bp in length
*	61311		gap	of unknown	length
*	61411	65807:	contig	of 4397	bp in length
*	65808		gap	of unknown	length
*	65908	70160:	contig	of 4253	bp in length
*	70161		gap	of unknown	length
*	70261	76919:	contig	of 6659	bp in length
*	76920		gap	of unknown	length
*	77020	85764:	contig	of 8745	bp in length
*	85765		gap	of unknown	length
*	85865	106630:	contig	of 20766	bp in length
*	106631		gap	of unknown	length
*	106731	134020:	contig	of 27290	bp in length
*	134021		gap	of unknown	length
*	134121	165082:	contig	of 30962	bp in length
*	165083		gap	of unknown	length
*	165103	196240:	contig	of 31058	bp in length
*	196241		gap	of unknown	length
*	196341	232180:	contig	of 35840	bp in length.
FEATURES	Location/Qualifiers				
SOURCE	1.. 232180				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/Clone="Rp11-588H15"				
misc_feature	1.. 1312				
	/note="assembly_name:Contig14"				
misc_feature	1413.. 2446				
	/note="assembly_name:Contig15"				
misc_feature	2547.. 3878				
	/note="assembly_name:Contig30"				
misc_feature	3979.. 5215				
	/note="assembly_name:Contig35"				
misc_feature	5316.. 6544				
	/note="assembly_name:Contig40"				
misc_feature	6645.. 7940				
	/note="assembly_name:Contig42"				
misc_feature	8041.. 9190				
	/note="assembly_name:Contig44"				
misc_feature	9291.. 10909				
	/note="assembly_name:Contig47"				

QY 1196 ACTGTGTGAAGAAAAGCA 1213  
 Db 178629 CTGTGTCATATTCAGTA 178612  
 RESULT 10  
 AC126290/c  
 DEFINITION Rattus norvegicus clone CH230-162N10, \*\*\* SEQUENCING IN PROGRESS  
 AC126290  
 VERSION 2 GI:21702826  
 KEYWORDS HTG; HTGS-PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 174143)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barbaccia,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,X., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newson,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
 Sodergren,E., Sonaikar,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 174143)  
 Worley,K.C.  
 Direct Submission  
 TITLE Submitted (05-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 174143)  
 Worley,K.C.  
 Direct Submission  
 TITLE Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA  
 On Jul 7, 2002 this sequence version replaced gi:21699099.  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GIBG  
 Center clone name: CH230-162N10  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 107819 bases at least Q40  
 Consensus quality: 114791 bases at least Q30  
 Consensus quality: 119520 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 70 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 1306: contig of 1306 bp in length  
 1307 1406: gap of unknown length  
 1407 2496: contig of 1090 bp in length  
 2497 2596: gap of unknown length  
 2597 3844: contig of 1248 bp in length  
 3845 3944: gap of unknown length  
 3945 5149: contig of 1204 bp in length  
 5149 5249: gap of unknown length  
 5249 6658: contig of 1409 bp in length  
 6658 6757: gap of unknown length  
 6757 7934: contig of 1177 bp in length  
 7935 8034: gap of unknown length  
 8034 9623: contig of 1589 bp in length  
 9623 9724: gap of unknown length  
 9724 11037: contig of 1314 bp in length  
 11037 11138: gap of unknown length  
 11138 12384: contig of 1247 bp in length  
 12385 12484: gap of unknown length  
 12485 13518: contig of 1034 bp in length  
 13519 13618: gap of unknown length  
 13619 14928: contig of 1310 bp in length  
 14929 15028: gap of unknown length  
 15029 15466: contig of 1518 bp in length  
 15467 16647: gap of unknown length  
 16647 17903: contig of 1257 bp in length  
 17904 18003: gap of unknown length  
 18004 19034: contig of 1031 bp in length  
 19035 19135: gap of unknown length  
 19135 20550: contig of 1416 bp in length  
 20551 21854: gap of unknown length  
 21855 21954: gap of unknown length  
 21955 23112: contig of 1358 bp in length  
 23113 23412: gap of unknown length  
 23413 24506: contig of 1094 bp in length  
 24507 24606: gap of unknown length  
 24607 26334: contig of 1728 bp in length  
 26335 26434: gap of unknown length  
 26435 27461: contig of 1027 bp in length  
 27462 27561: gap of unknown length  
 27562 29384: contig of 1822 bp in length  
 29384 31029: gap of unknown length  
 31030 31129: gap of unknown length  
 31130 32142: contig of 1013 bp in length  
 32142: contig of 1013 bp in length

[illegible]



Consensus quality: 119520 bases at least Q20

\*\*\*\*\*  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 70 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 1306: contig of 1306 bp in length  
\* 1307 1406: gap of unknown length  
\* 1407 2496: contig of 1090 bp in length  
\* 2497 2596: gap of unknown length  
\* 2597 3844: contig of 1248 bp in length  
\* 3845 3945: gap of unknown length  
\* 3946 5149: contig of 1204 bp in length  
\* 5149 5248: gap of unknown length  
\* 5249 6657: contig of 1409 bp in length  
\* 6658 6757: gap of unknown length  
\* 6758 7934: contig of 1177 bp in length  
\* 7935 8035: gap of unknown length  
\* 8036 9624: contig of 1589 bp in length  
\* 9625 9724: gap of unknown length  
\* 9725 11037: contig of 1314 bp in length  
\* 11038 12384: contig of 1247 bp in length  
\* 12385 12485: gap of unknown length  
\* 12486 13518: contig of 1034 bp in length  
\* 13519 13618: gap of unknown length  
\* 13619 14928: contig of 1310 bp in length  
\* 14929 15028: gap of unknown length  
\* 15029 16546: contig of 1518 bp in length  
\* 16547 16647: gap of unknown length  
\* 16648 17903: contig of 1257 bp in length  
\* 17904 18003: gap of unknown length  
\* 18004 19034: contig of 1031 bp in length  
\* 19035 19135: gap of unknown length  
\* 19136 20550: contig of 1416 bp in length  
\* 20551 20650: gap of unknown length  
\* 20651 21854: contig of 1204 bp in length  
\* 21855 21954: gap of unknown length  
\* 21955 23312: contig of 1358 bp in length  
\* 23313 23412: gap of unknown length  
\* 23413 24506: contig of 1094 bp in length  
\* 24507 24606: gap of unknown length  
\* 24607 28334: contig of 1728 bp in length  
\* 28335 28434: gap of unknown length  
\* 28435 27461: contig of 1027 bp in length  
\* 27462 27561: gap of unknown length  
\* 27562 29383: contig of 1822 bp in length  
\* 29384 29484: gap of unknown length  
\* 29485 31029: contig of 1546 bp in length  
\* 31030 31129: gap of unknown length  
\* 31130 32142: contig of 1013 bp in length  
\* 32143 32432: gap of unknown length  
\* 32433 33607: contig of 1365 bp in length  
\* 33608 33707: gap of unknown length  
\* 33708 34764: contig of 1057 bp in length  
\* 34765 34864: gap of unknown length  
\* 34865 37347: contig of 2483 bp in length  
\* 37348 37447: gap of unknown length  
\* 37448 38924: contig of 1477 bp in length  
\* 38925 39024: gap of unknown length  
\* 39025 40485: contig of 1461 bp in length  
\* 40486 40585: gap of unknown length  
\* 40586 41695: contig of 1110 bp in length  
\* 41696 41795: gap of unknown length  
\* 41796 44470: contig of 2675 bp in length  
\* 44471 44570: gap of unknown length  
\* 44571 45985: contig of 1415 bp in length

\* 45986 46085: gap of unknown length  
\* 46086 47391: contig of 1306 bp in length  
\* 47392 47491: gap of unknown length  
\* 47492 49482: contig of 1991 bp in length  
\* 49483 49582: gap of unknown length  
\* 49583 51236: contig of 1654 bp in length  
\* 51237 51336: gap of unknown length  
\* 51337 52989: contig of 1653 bp in length  
\* 52990 53089: gap of unknown length  
\* 53090 54559: contig of 1470 bp in length  
\* 54560 54659: gap of unknown length  
\* 54660 56814: contig of 2155 bp in length  
\* 56815 56914: gap of unknown length  
\* 56915 58318: contig of 1404 bp in length  
\* 58319 58418: gap of unknown length  
\* 58419 59961: contig of 1543 bp in length  
\* 59962 60061: gap of unknown length  
\* 60062 62529: contig of 2468 bp in length  
\* 62530 62629: gap of unknown length  
\* 62630 64062: contig of 1433 bp in length  
\* 64063 64162: gap of unknown length  
\* 64163 66181: contig of 2019 bp in length  
\* 66182 66281: gap of unknown length  
\* 66282 69506: contig of 3225 bp in length  
\* 69507 69606: gap of unknown length  
\* 69607 72241: contig of 2635 bp in length  
\* 72242 72341: gap of unknown length  
\* 72342 74220: contig of 1879 bp in length  
\* 74221 74320: gap of unknown length  
\* 74321 77101: contig of 2781 bp in length  
\* 77102 77201: gap of unknown length  
\* 77202 79049: contig of 1848 bp in length  
\* 79050 79149: gap of unknown length  
\* 79150 81909: contig of 2760 bp in length  
\* 81910 82009: gap of unknown length  
\* 82010 84999: contig of 2990 bp in length  
\* 85000 85099: gap of unknown length  
\* 85100 87219: contig of 2119 bp in length  
\* 87219 87318: gap of unknown length  
\* 87319 90332: contig of 3014 bp in length  
\* 90333 90432: gap of unknown length  
\* 90433 92429: contig of 1997 bp in length  
\* 92430 92529: gap of unknown length  
\* 92530 95472: contig of 2943 bp in length  
\* 95473 95572: gap of unknown length  
\* 95573 98072: contig of 2500 bp in length

Query Match 10.8%; Score 158; DB 2; Length 174143;  
Best Local Similarity 91.8%; Pred. No. 4.1e-29;  
Matches 178; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1008 GAGGAGAGTGTCCACC-GTCCGAGTCCCGACTCATGAGGAGGATCATACCTATTTGG 1066  
Db 163183 GCGGAGAAGTAGTCCCGGTTTCGAGTCCCTCATGAGGAGGATCATACCTATTTGG 163242  
QY 1067 GAATTTGCCACAGACAGTTATGACATTTGGGTTTGGGTTTATTTGAATGACAGACTCT 1126  
Db 163243 GAATTTGCCACAGACAGTTATGACATTTGGGTTTGGGTTTATTTGAATGACAGACTCT 163302  
QY 1127 CCAATATGCTGTCTGACGTGTCATGTGTCAGTGCAGTGCAGTGCAGGAGGAGGAGGAGAA 1186  
Db 163303 CCAAAACGCTGTCTGACGTGTCATGTGTCAGTGCAGTGCAGGAGGAGGAGGAGAA 163362  
QY 1187 GAAATGTCTACTTG 1200  
Db 163363 GAAGGTAGGCCTTG 163376

RESULT 14

AK095650  
LOCUS AK095650 3547 bp mRNA linear PRI 15-JUL-2002  
DEFINITION Homo sapiens cDNA FLJ38331 fis, clone FCBBF3025285, moderately  
similar to Mus musculus peripheral benzodiazepine receptor

```
associated protein (Pap7) mRNA.
ACCESSION AK095650
VERSION AK095650.1 GI:21754954
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib.FCBBF3
clone:FCBBF3025285.

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Kanehori,K., Ishibashi,T., Chiba,Y., Fujimori,K., Hiraoka,S.,
Tanai,H., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Watanabe,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,
Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A., Oshima,A.,
Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K.,
Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3547)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES             source
    source
        location/Qualifiers
            1..3547
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="FCBBF3025285"
                /tissue_type="brain"
                /clone_lib="FCBBF3"
                /dev_stage="fetal"
                /note="cloning vector: pME18SFL3"
BASE COUNT   928 a   768 c   832 g   1019 t
ORIGIN
Query Match      10.3%; Score 150.8; DB 9; Length 3547;
Best Local Similarity 57.7%; Pred. No. 2.le-27;
Matches 306; Conservative 0; Mismatches 197; Indels 27; Gaps 1;

QY 899 AAAGACTCTCTCCAGTGATTGCGAGCTCCATCCATGCGAAGACACCAATCAAGAC 958
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 495 AAAGTCTCCACCTCTGATGGCTCCTCCATGCACTCGAGCCTTTGCCAAGGTGAAGAA 554
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 959 TTAAAGAAAGATTCGGCAGGATGCGAGTCTGTGATTACAGATCGTCGAGGAGAGTC 1018
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1019 GTACCCGTCCAGTCCCGACATCAGGAAGATCATACCTATTTTGGGAATTTGCCACA 1078
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 615 GTGACCATCCGGGTACCTACTCCTCCAGAGGGGAGCGTGTCTGCTGGGAGTTTGGCACC 674
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1079 GACAGTTATGACATGGGTTTGGGTTTATTTTGAATGACAGACTCTCCAAATGCTGCT 1138
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 675 GATGACTATGACATGGCTTTGGAGTTTATTTTACTGGACCCCTGTAACTAGCAGCTGAC 734
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1139 GTCAGTGTGCATGTGTCAGTGTCCAGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1195
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 735 ATAAGTGTGCAGGTGAGTATCCAGTACCATGAGGATGAAGAGGAGGAGGAGGAGGAG 794
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1196 -----ACTTGTGAAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 795 GAGATTGAAGAACCCGTTCCAGCTGGAGATGTGGAGAGAGGCTCCAGGAGGCTCTTGGCG 854
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

QY 1232 CCTCTGCTGGATGAGATTGTACCTGTGTACCGGGGGAGTCTACAGGAGAGATATATGCA 1291
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 855 GGTGCTATGGGAGGTTCATCCCTGTGTACCGGGGGAGACCCAGGAGAGGTGCAGGCT 914
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1292 GGCAGCCACCATATCCAGGAGGGAGTCTATCTCCTCAAGTTTGTAAATTCCTACTCT 1351
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 915 GGCAGCCATGACTACCTCGTGGTGAGGCGATCTACCTGCTCAAGTTTCGACAACTCTCTACTCC 974
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1352 CTGTGGAGGTCCAGTCCGTCTACTACAGATCTATATATAGATAGAG 1401
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 975 CTGCTGGCAACAAGAGCTCTCTACTTCCACATCTACTACACAGCTGAAG 1024
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

RESULT 15
AX400067 AX400067 882 bp DNA linear PAT 06-JUN-2002
LOCUS Sequence 238 from Patent WO0218424.
ACCESSION AX400067
VERSION AX400067.1 GI:21336485
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Tang,Y.T., Asundi,V., Zhou,P., Xue,A.J., Ren,F., Zhang,J.,
Wang,J.R., Zhao,Q.A., Wang,D., Liu,C., Drmanac,R.T. and Wehrman,T.
Nucleic acids and polypeptides
Patent: WO 0218424-A 238 07-MAR-2002;
HYSEQ, INC. (US)

FEATURES             source
    source
        location/Qualifiers
            1..882
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAD33511.1"
                /db_xref="GI:21336486"
                /translation="MVSPVSKDATEDLRKATGPLEAQLVKDQLLPADQAVLNEMAK
                YOVPQSGDIVMIQSEHTGAIDVLSADLESADLGDHRKVPPLMAPPCITWFAKVE
                FSKLQKRNRLVVRGEVTVIRVPTHPGKRVCFWFDATDDIDFGVTFDWTPTVTS
                TDITVOVSDSDDEDEEEETEPVPAGDVVERSRSLRGRYGEVMPVYRRDSHR
                DVOAGSHDYPGEGIYLLKFDNSYSLRNKTLVFIHYTS"
BASE COUNT   221 a   218 c   250 g   193 t
ORIGIN
Query Match      10.3%; Score 150; DB 6; Length 882;
Best Local Similarity 57.8%; Pred. No. 2.9e-27;
Matches 304; Conservative 0; Mismatches 195; Indels 27; Gaps 1;

QY 899 AAAGACTCTCTTCCAGTGATTGCGAGTCCATCCATGCGAAGACACCAATCAAGAC 958
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 355 AAAGTCTCCACCTCTGATGGCTCCTCCATGCACTCGAGCCTTTGCCAAGGTGAAGAA 414
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 959 TTAAAGAAAGATTCGGCAGGATGCGAGTCTGTGATTACAGTACGTCGAGGAGAGTC 1018
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 415 TTCAAAGCAAGCTGGGCAAGAGAAAGACAGCCGCTGTGGTGGTGAAGCGTGGTGGAGTG 474
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1019 GTCACCGTCCGAGTCCCGACATCATGAGGAAGATCATACCTATTTTGGGAATTTGCCACA 1078
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 475 GTGACCATCCGGGTACCTACTCCTCCAGAGGGGAGCGTGTCTGCTGGAGTTTGGCACC 534
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1079 GACAGTTATGACATGGGTTTGGGTTTATTTTGAATGACAGACTCTTCCAAATGCTGCT 1138
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 535 GATGACTATGACATGGCTTTGGAGTTTATTTTACTGGACCCCTGTAACTAGCAGCTGAC 594
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1139 GTCAGTGTGCATGTGTCAGTGTCCAGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1193
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 595 ATAAGTGTGCAGGTGAGTATCCAGTACCATGAGGATGAAGAGAGAGAGAGAGAGAGAA 654
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY 1194 -----TCACTTGTGAAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
```



Db 655 GAGATTGAAGAACCGTTCCAGCTGGAGATGTGGAGAGAGGCTCCAGGAGCTCCTTGCGG 714  
QY 1232 CCTCTGCTGGATGAGATTGTACCTGTGTACCGCGGGACTGTACAGGAGGAGTATATGCA 1291  
Db 715 GGTGCTATGGGGAGGTGATGCTGTGTACCGCGGGAGACGCCACCGAGACGTGCAGGCT 774  
QY 1292 GGCAGCCACCACTATCCAGGGAGGGAGTCTATCTCTCAAGTTTGATAATTCCTACTCT 1351  
Db 775 GGCAGCCATGACTACCTGGTGAGGGCATCTACCTGCTCAAGTTCGACAACCTCTACTCC 834  
QY 1352 CTGTGGAGTCCAAAGTCCCTCTACTACAGAGTCTATTATTACTAGAT 1397  
Db 835 CTGCTGGCAACAAGACTCTCTACTTCCACATCTACTACACAGCT 880

Search completed: February 21, 2003, 23:10:38  
Job time : 5171 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**